h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES
fbiuw
FACHBEREICH INFORMATIONS-
UND WISSENSMANAGEMENT

Schlagwortvergabe durch Nutzende (Tagging) als Hilfsmittel zur Suche im Web.

Ansatz, Modelle, Realisierungen

Bearbeiter: Sascha Carlin, 586773

Referent: Prof. Dr. Reginald Ferber

Korreferent: Prof. Dr. Christian Otto

Abgabe: 14.08.2006.

Erklärung

Ich versichere, dass ich die vorliegende Diplomarbeit selbständig erarbeitet habe und dass dabei keine anderen als die angegebenen Quellen und Hilfsmittel verwendet wurden.

Ort, Datum	Unterschrift

Erklärung

Mit der Ausleihe der Diplomarbeit sind wir einverstanden.

Student

Ort, Datum	Unterschrift

Betreuender Professor

Ort, Datum	Unterschrift

Inhaltsverzeichnis

Individuals and interactions

over processes and tools

Responding to change

over following a plan

Aus dem Agile Manifesto for Software Development

As long as we don't let the ontologists take over and tell us why tags are all wrong, need to be classified into domains, and need to be systematized, this is going to work well albeit, sloppily.

Adam Bosworth, Vice President of Engineering at Google Inc.

Gewidmet meinen Eltern für mein Leben.

1 Einleitung

Nach dem zu Beginn der Ära des World Wide Web von Hand gepflegte Linklisten und -Verzeichnisse und an Freunde und Kollegen per E-Mail verschickte Links genügten, um die Informationen zu finden, nach denen man suchte, waren schon bald Volltextsuchmaschinen und halbautomatisch betriebene Kataloge notwendig, um den mehr und mehr anschwellenden Informationsfluten des Web Herr zu werden.

Heute bereits sind diese Dämme gebrochen und viele Millionen Websites halten Billionen an Einzelseiten mit Informationen vor, von Datenbanken und anderweitig versteckten Informationen ganz zu schweigen. Mit Volltextsuchmaschinen erreicht man bei dieser Masse keine befriedigenden Ergebnisse mehr. Entweder man erzeugt lange Suchterme mit vielen Ausschließungen und ebenso vielen nicht-exklusiven ODER-Verknüpfungen um verschiedene Schreibweisen für den gleichen Term abzudecken oder man wählt von vornherein die Daten-Quelle, an die man seine Fragen stellt, genau aus. Doch oft bleiben nur klassische Web-Suchmaschinen übrig, zumal wenn der Fragende kein Informationsspezialist mit Kenntnissen von Spezialdatenbanken ist, sondern, von dieser Warte aus gesehenen, ein Laie.

Und nicht nur im Web selbst, auch in unternehmensinternen Intranets steht man vor diesem Problem. Tausende von indizierten Dokumente mögen ein Eckdatum sein, nach dem sich der Erfolg der Einführung eines Intranets bemessen lässt, aber eine Aussage über die Nützlichkeit ist damit nicht getroffen. Und die bleibt meist hinter den Erwartungen zurück, vor allem bei denen Mitarbeitern, die tatsächlich mit dem Intranet arbeiten müssen.

Entscheidend ist für die Informationsauffindung in Inter- und Intranet eine einfach zu nutzende und leicht anpassbare Möglichkeit, neue interessante Inhalte zu entdecken.

Mit Tags steht eine mögliche Lösung bereit.

a Taggingsysteme

i Grundsätzliches

Ein auf Tags basierendes Informationssystem gibt den Nutzern die Möglichkeit, eigene oder fremde Dokumente im System bekannt zu machen und mit einem oder mehreren Tags zu versehen. Je nach System müssen die Nutzer dabei keine oder nur wenige formale Daten erschließen, wobei es auch Systeme gibt, die dies verlangen. Die Eingabe der zwingend benötigten Daten soll für die Nutzer so einfach wie möglich gestaltet sein. Je nach Ausrichtung des Systems (siehe Kapitel Ausprägungen von Taggingsystemen, S. 30ff) erlaubt das System den Nutzern, eigene Sammlungen (auch „Bibliotheken", „Listen" genannt) anzulegen und in ihnen nach vergebenen Tags zu navigieren bzw. zu filtern. Solche Systeme werden zur Verwaltung von Bookmarks (Favoriten) oder Bildern benutzt, zur Pflege von Literaturlisten und Bibliographien, oder zum Auffinden und Bewerten von Nachrichten, Restaurants und ähnlichem.

ii del.icio.us

del.icio.us, 2003 von Josua Schachter gegründet[1], ist eine Bookmark-Verwaltung. Anders als die aus Web-Browsern bekannten Ansätze speichert del.icio.us die gesammelten Verweise nicht auf einem Personal Computer, sondern zentralisiert auf der del.icio.us-Website[2]. Somit haben die Nutzer von jedem Internetzugang aus Zugriff auf ihre Bookmarks.

Bookmark-Verwaltungen in Web-Browsern orientieren sich in ihrer Struktur an Dateiverwaltungssystemen mit Verzeichnissen (Ordern), und damit an einem hierarchischen Ansatz. Unterhalb des „Hauptverzeichnisses" kann der Nutzer Verweise auf Webseiten (Bookmarks) anlegen, Unterverzeichnisse erstellen und so eine Hierarchie erzeugen, in der er seine Bookmarks ablegen kann.

1 Siehe http://del.icio.us/about/info – Zugegriffen am 19.06.2006.
2 Siehe http://del.icio.us/

Nachteilig dabei ist zum einen, dass eine solche Bookmark-Sammlung nur auf dem eigenen Personal Computer benutzbar ist, zum anderen dass sie nur mit dem gleichen Browser benutzt werden kann, mit dem sie erstellt wurde.

Auch die hierarchische Gliederung bringt Nachteile mit sich. So kann ein Bookmark immer nur einer Klasse (Ordner) zugeordnet werden, nicht aber mehreren (es sei denn, der Benutzer kopiert ihn in jeden Ordner, in dem er erscheinen soll).

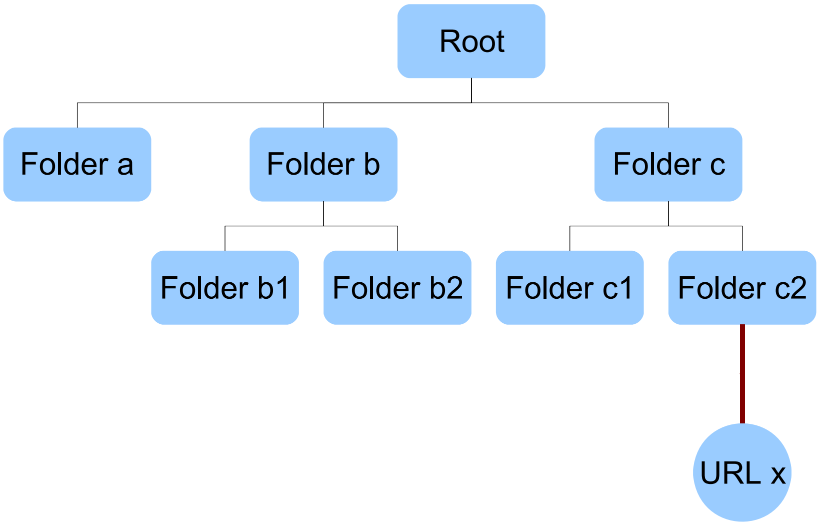


Abbildung 1-1: Schematische Darstellung der hierarchischen Struktur von Bookmark-Sammlungen in Webbrowsern Die Ressource „URL x“ ist in Ordner c2 abgelegt.

Abgesehen davon bieten die Bookmark-Verwaltungen der Web-Browser nur sehr rudimentäre Möglichkeiten der Suche und Verwaltung. Zumeist beschränken sich die Möglichkeiten auf das Anlegen, Verändern und Löschen von Bookmarks und Ordnern.

del.icio.us bietet weitaus mehr. Als Werkzeug für Schachters persönliche Bookmark-Sammlung geschrieben (genannt „muxway“, vgl. [Norton2006]), ermöglicht del.icio.us die Ablage von Bookmarks unter beliebigen und beliebig vielen Schlagworten: Tags.

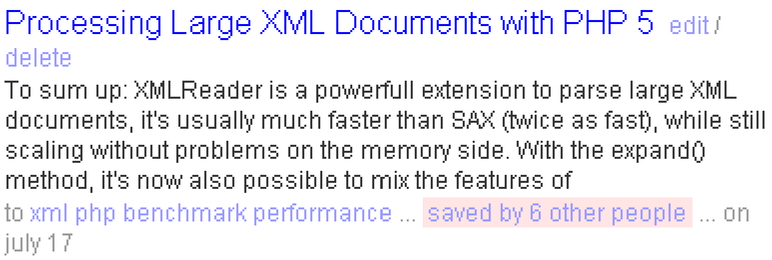


Abbildung 1-2: Ein einzelner Bookmark in del.icio.us, wie ihn ein Nutzer in seiner eigenen Bookmarks-Sammlung sieht. Zu sehen ist der Titel der verlinkten Webseite, ein Zitat aus der Seite, die benutzten Tags, die Anzahl der übrigen Nutzer, die diese Webseite ebenfalls gebookmarkt haben sowie das Datum, an dem die Webseite gebookmarkt wurde.

Abbildung 1-2 zeigt einen Bookmark aus der Bookmark-Sammlung eines del.icio.us-Nutzers. Dieser Bookmark wurde mit den Tags „xml“, „php“, „benchmark“ und „performance“ getaggt. Sobald ein Nutzer ein Tag vergibt, erzeugt er damit einen Pseudo-Ordner (auch „bucket“ genannt). Der in Abbildung 1-2 gezeigte Bookmark ist in den Pseudo-Ordnern „xml“, „php“, „benchmark“ und „performance“ abgelegt.

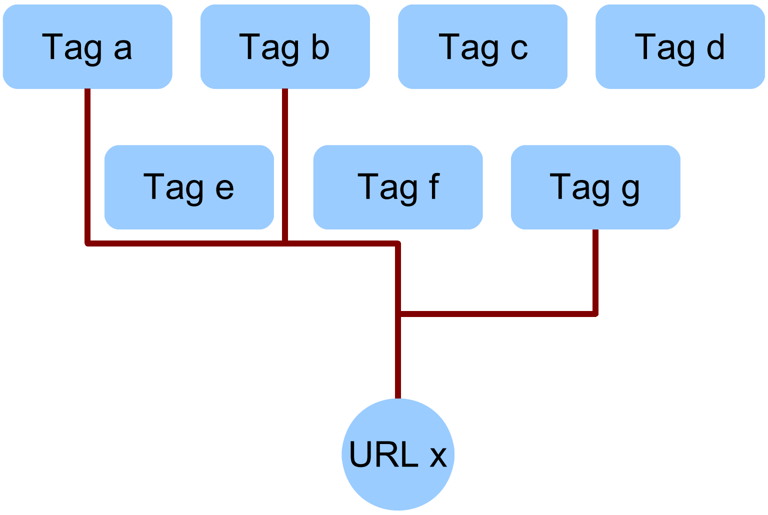


Abbildung 1-3: In Bookmark-Verwaltungen wie del.icio.us, die auf Tags basieren, kann ein Bookmark zu mehreren Tags - "Ordnern" – gehören. Die Ressource „URL x“ ist in den „Ordnern“ Tag a, Tab b und Tag g abgelegt.

Auch die Suche bzw. das Browsen in der Bookmark-Sammlung wird von del.icio.us stark vereinfacht. Im Kapitel Taggingsysteme am Beispiel von del.icio.us, S. 63ff, wird dies deutlich.

iii Flickr

Mit der Online-Bildverwaltung Flickr[3] steht ein Dienst bereit, mit dem Fotografien, Bilder und Illustrationen jeder Art online verwaltet werden können. Wie del.icio.us ermöglicht Flickr dabei die Eingabe von Tags zur Indexierung der Ressource. Im Gegensatz zu del.icio.us, wo jede Ressource in den Bookmark-Sammlungen mehrerer Nutzer auftauchen und daher von mehreren Menschen getaggt werden kann, erlaubt Flickr nur demjenigen Nutzer, der eine Ressource erstellt, sie zu taggen. In der Literatur wird hierbei nach „set-model" und „bag-model" unterschieden (siehe Aggregation, Seite 35).

iv CiteULike

Wo del.icio.us eine Anwendung zur Sammlung generischer Bookmarks – oder besser: per HTTP erreichbarer Ressourcen – ist, sieht sich CiteULike als Dienst für wissenschaftlich arbeitende Menschen und fokussiert daher auf wissenschaftliche Ressourcen wie Zeitschriftenaufsätze, Proceedings usw. Dazu stellt[4] es eigens entwickelte Schnittstellen zu Systemen wie ACM Digital Library[5], CiteSeer[6] oder IEEE Xplorer[7] bereit, mit denen die Übernahme bibliographischer Daten weitgehend automatisiert wird. Darüber hinaus ist auch CiteULike tagbasiert und ermöglicht neben der Suche nach Autoren und Titel die Filterung nach Tags. Als vor allem für wissenschaftlich arbeitende Menschen erstelltes Werkzeug bietet CiteULike die Möglichkeit der Eingabe bibliographischer Daten im Rahmen einer Formalerschließung. Darüber hinaus gibt es Im- und Exporter für BibTex und EndNote, Formate zur Speicherung bibliographischer Daten für Literaturverwaltungssoftware.

Im Kapitel Kollaborativer Ansatz, S. 38ff, beschreibe ich ein System, das anhand CiteULike zeigt, wie Tagging bzw. Folksonomies für Collaborative Filtering eingesetzt werden können.

3 Siehe http://www.flickr.com/ - Zugegriffen am 19.07.2006.
4 Siehe http://www.citeulike.org/faq/all.adp – Zugegriffen am 19.07.2006.
5 Siehe http://portal.acm.org/dl.cfm – Zugegriffen am 19.07.2006.
6 Siehe http://citeseer.ist.psu.edu/ - Zugegriffen am 19.07.2006.
7 Siehe http://ieeexplore.ieee.org/ - Zugegriffen am 19.07.2006.

v Weitere Systeme

Hammond et al. (vgl. [Hammond2005]) betrachten in ihrer Übersicht über bestehende Taggingsysteme neun Systeme: die bereits erwähnten del.icio.us, CiteULike und Flickr, und darüber hinaus Connotea, Frassle, Furl, Simpy, Spurl.net sowie unalog. Weiterhin erwähnenswert sind zum Beispiel steve.museum und Qype, mit denen im Gegensatz zu den vorgenannten Ausstellungsstücke in Museen sowie Örtlichkeiten getaggt werden können.

steve.museum

Getragen von mehreren namhaften nordamerikanischen Museen, zum Beispiel dem Guggenheim Museum oder dem The Metropolitan Museum of Art, ist es das Ziel dieser Initiative[8], Museumsbesuchern die Möglichkeit zu geben, Ausstellungsstücke mit Tags zu beschreiben. Die Beweggründe lauten[9]:

> *People see different things. What a specialist notices may not be important to a member of the general public, and vice versa, and the words that a curator uses may not be familiar to the average museum-goer. Allowing users [to] describe collections—using their own vernacular or language—may help other users find things that interest them.*

Man sieht die Vorteile unter anderem in einer engeren Verbindung zwischen Museumsbesuchern und den Museen[10]:

> *User-based description also allows people to develop a more personal relationship to museums and museum collections. They can contribute their knowledge, and see the improvements that result from sharing their expertise.*

8 Siehe http://www.steve.museum/ - Zugegriffen am 24.06.2006.
9 Siehe http://www.steve.museum/index.php?option=com_content&task=view&id=38&Itemid=61 – Zugegriffen am 24.06.2006.
10 Siehe http://www.steve.museum/index.php?option=com_content&task=view&id=42&Itemid=61 – Zugegriffen am 24.06.2006.

In einer Veröffentlichung vom März 2006 (vgl.. [Chun2006]) heißt es weiter:

> *Museums have an opportunity to enhance user experience and to experiment with new strategies for user engagement that builds community and attract new audiences.*

Mit der Hilfe von Freiwilligen sollen ab September 2006 Objekte aus den Sammlungen der teilnehmenden Museen getaggt werden. Dazu werden Web-Kataloge mit Fotografien der Objekte bereitgestellt. Die teilnehmenden Freiwilligen können dann in diesen Kataloge browsen und die Objekte taggen.

Powerhouse Museum Collection

Das Powerhouse Museum in Sydney ist nach eigenen Angaben das größte und populärste Museum Australiens und beherbergt 385.000 Objekte verschiedenster Art[11]. Auch hier werden von Freiwilligen Objekte mit Tags indexiert. Der Ansatz des Powerhouse Museum verknüpft bei der Suche auf der Website die Tags mit dem Thesaurus, nach dem die Objekte des Museums im Katalog erschlossen sind.



Abbildung 1-4: Ausschnitt aus dem Suchergebnis für "bowl" in der Sammlung des Powerhouse Museum. Oben die von Benutzern eingegeben Tags, die mit "bowl" in Beziehung stehen, unten die Klassen aus dem Thesaurus des Museums.

Qype

Bei Qype handelt es sich um ein Tagging-System für Örtlichkeiten wie Bars und Restaurants, aber auch für Geschäfte, Handwerker und jede Art von lokaler Dienstleistung. Die Be-

11 Siehe http://www.powerhousemuseum.com/about/in.dex.asp – Zugegriffen am 06.08.2006.

nutzer können, nach Städten geordnet, Informationen zu den diversen Dienstleistungen abrufen , sie rezensieren und ihnen Tags und Bewertungen zuweisen. Die Suche von Qype basiert hauptsächlich auf den Tags, wobei aber auch nach formalen Angaben wie Firmennamen und ähnlichem gesucht werden kann. Die Implementierung erinnert an das Recommendation-System von Amazon.com, das in der US-amerikanischen Version seit einiger Zeit ebenfalls Tagging unterstützt[12].



Abbildung 1-5: Suchergebnis von Qype für "bücher" in Darmstadt. Es zeigt neben den Treffern links auf der rechten Seite eine Karte mit den Standorten der Treffer. Außerdem werden zu "bücher" verwandte Tags angezeigt.

12 Siehe http://www.amazon.com/gp/product/0596007655/ - Auf dieser Produktseite sind neben Rezensionen und Bewertungen auch die Tags zu sehen, mit denen Amazon-Kunden das Buch „Ambient Findability“ getaggt haben. Zugegriffen am 08.08.2006.

2 Tags – freie Schlagwörter vergeben durch Nutzende

Im Folgenden wird anhand von Beispielen gezeigt, welche Arten von Tags es gibt und welche Probleme durch die freie Verschlagwortung mit Tags entstehen. Es wird beschrieben, aus welchen Motivationen heraus Nutzer Tags vergeben. Weiter wird gezeigt, anhand welcher Kriterien man Taggingsysteme vergleichen kann. Dann werden Verbesserungsvorschläge gemacht, wie der Taggingprozess selbst verbessert werden könnte, um die beschriebenen Probleme teilweise zu lösen. Den Schluss dieses Kapitels bildet die Erläuterung eines Collaborative Filtering-Systems anhand von CiteULike.

a Definition

Der Begriff **tag** stammt aus der englischen Sprache und meint eine Markierung in Form eines Stück Papiers oder Pappe, dass an einem Gegenstand angebracht wird, um diesen zu markieren (vgl. Dic1983]):

> *a small narrow piece of paper, material, etc., fixed to something to show what it is, who owns it, details about it, etc.*

Diese Definition aus einem einsprachigen Wörterbuch zeigt bereits, dass ein Tag benutzt wird, um Informationen über einen Gegenstand, also Metadaten, aufzunehmen bzw. zu transportieren. Weiterhin lässt diese Definition offen, wie viele Tags einem Gegenstand zugeordnet werden und von wem. Es ist also ohne weiteres möglich, einem Gegenstand mehrere Tags zuzuordnen, und es ist auch möglich, dass mehrere Personen dies unabhängig von einander tun. Auch legt die Definition nicht fest, welcher Art die Informationen, die einem Gegenstand so zugeordnet werden, sein sollten. Fest steht, dass es Informationen sind, die den Gegenstand näher betrachten und ihn – zumindest teilweise – beschreiben.

Tags können Aufschluss geben über verschiedene Facetten bzw. Eigenschaften eines Gegenstands. Golder und Huberman (vgl. [Golder2005]) nennen sieben Arten von Tags:

1. Identifying what (or who) it is about
2. Identifying what it is
3. Identifying who owns it
4. Refining categories
5. Identifying qualities or characteristics
6. Self reference
7. Task organizing

Zum dritten Punkt, *who owns it*, lässt sich hinzufügen *who created it*. So zumindest meinen Golder und Huberman diesen Punkt, wenn sie von *content ownership* sprechen.

An dieser Einteilung wird klar, dass Tags mehr sind als freie Schlagworte oder Deskriptoren, die sich ausschließlich auf den Inhalt und das Themengebiet des Inhalts eines Gegenstands beziehen. dass Tags nicht aus einem kontrollierten Vokabular stammen, und somit auch keine Schlagworte im klassischen Sinn sein können, erschließt sich aus der Bandbreite der möglichen Arten. Sie sind auch mehr als Stichwörter, die aus formalen Angaben zu einem Gegenstand gewonnen werden.

Dazu kommt, dass Tags im Gegensatz zu Schlag- und Stichwörtern weder in automatischen Verfahren noch von geschulten Informationsspezialisten vergeben werden, sondern von den Endnutzern der Datensammlungen selbst.

In dieser Arbeit meint der bis hierher gebrauchte Begriff *Gegenstand* natürlich *Dokumentarische Beschreibungseinheit* (DBE). Diese Dokumente sind, wenn von Tags die Rede ist, direkt per HTTP adressierbare Dokumente wie Webseiten, Bilder und andere Mediendateien. An-

stelle von Gegenstand oder Dokument wird in dieser Arbeit auch von Ressourcen gesprochen.

Kommen wir nun noch einmal zu den Klassen von Tags nach Golder und Huberman zurück.

i Identifying what (or who) it is about

Solche Tags nehmen direkten Bezug auf die in dem Dokument diskutierten Themengebiete oder Personen und Organisationen. Diese Arbeit könnte beispielsweise mit den folgenden Tags versehen werden: *„tags“, „tagging“, „anwendungen“, „folksonomy“, „taxonomien“, „übersicht“, „collaborative-filtering“.*

In den meisten Fällen sind diese Tags Nomen.

Anhand dieser Tags wird der Inhalt dieser Arbeit beschrieben, ohne auf formale Aspekte wie Autor oder Dokumenttyp einzugehen . Zu diesen Eigenschaften kommen wir nun.

ii Identifying what it is

Ohne auf den Inhalt einzugehen, beschreiben Tags dieser Klasse, was für ein Dokument hier vorliegt. Für diese Arbeit wäre das passende Tag „diplomarbeit“.

Mit diesem Tag wird klar, dass es sich hier um eine Diplomarbeit handelt.

iii Identifying who it owns

Wie oben erläutert, nimmt man mit solchen Tags Bezug auf den Ersteller eines Dokuments. Dies kann eine Person oder eine Personengruppe sein oder eine Organisation. Beispiele: „sascha-carlin“, „fbiuw“.

iv Refining categories

Mit diesen Tags, die für sich allein genommen keine genaue Aussagekraft besitzen, können einzelne Tags verfeinert werden, um detaillierter auf eine Facette des Dokuments einzugehen. Zum Beispiel könnte das Tag „diplomarbeit“ mit dem Tag „zur erlangung des diploms im studiengang informations- und wissensmanagement“ verfeinert werden, um diese spezielle Eigenschaft des Dokuments aufzuzeigen. Kürzere Tags wären sicherlich sinnvoller, zum Beispiel „fbiuw“, „iuw“ oder „informationswirt“.

v Identifying qualities or characteristics

Tags dieser Klasse beschreiben den Dokumentinhalt durch passende Adjektive. Je nach Dokument macht es mehr oder weniger Sinn, solche Tags zu benutzen. Beispiele: „witzig“, „lang“, „informativ“, „lesenswert“, „schlecht“, „gut“.

vi Self reference

Dies sind Tags, die der Nutzer verwendet, um seine Beziehung zu dem Dokument oder dem Inhalt darzustellen. Kunstwörter, bestehend aus Possessivpronomen und Angaben zum Typ des Dokuments werden hier benutzt, zum Beispiel „meinearbeiten“ oder „meinemeinung“.

vii Task organizing

Damit gemeint sind Tags, die der Tagger benutzt, um Aufgaben mit einem Dokument zu verknüpfen, wie „lesen“ oder „ausdrucken“. In der englischen Sprache sind Tags wie „to-read“ (lies: to read) möglich.

Es ist erkennbar, das lediglich die ersten drei bis vier Klassen von Tags unabhängig vom Nutzer sein können. Mit den Tags der letzten zwei bis drei Klassen kann niemand außer demjenigen Nutzer, der sie eingab, für sich selbst etwas anfangen; sie sind intrinsisch in Relation zu der Person, die sie eingab, wohingegen die ersten drei Klassen extrinsisch sind.

b Probleme

Da jedes Wort jeder natürlichen Sprache – sowie Abkürzungen, Produktnamen und so weiter – zu einem Tag werden kann, entstehen ähnliche Probleme, die auch natürliche Sprachen haben, zum Beispiel in den Bereichen der Rechtschreibung (Syntax) und der Disambiguierung (Semantik). Da die meisten Taggingsysteme über das World Wide Web weltweit zur Verfügung stehen, entstehen natürlich auch Sprachbarrieren – wiederum in Syntax und Semantik. Letztlich werden die Problemfelder Benutzung von Sonderzeichen und Subjektivität von Informationen vorgestellt.

i Homonyme und Synonyme

Was ist mit dem Begriff „Dose“ gemeint? Eine Schachtel oder Box aus Blech, deren Deckel sich vollständig vom Unterteil lösen lässt – oder ein Personal Computer, auf dem ein Microsoft DOS-Betriebssystem läuft? Rein grammatikalisch lässt sich auch nicht sagen, ob mit „Froschgift“ ein Gift gegen Frösche (wie in „Rattengift“) gemeint ist oder die Umkehrung, Gift, das von Fröschen gewonnen wird.

Nutzen wir einen Tag wie „dose“, um – zugegebenermaßen satirische oder zumindest lustige – Informationen über Windows-PCs zu finden, müssten wir dem initialen Suchbegriff weitere hinzufügen, um disambiguieren zu können, weil uns im Moment keine andere Art Dose interessiert.

ii Sprachbarrieren

Die beiden Tags „deutsch“ und „german“ sind gleichbedeutend, wenn auch das eine Tag in deutscher Sprache und das zweite in der englischen geschrieben wurde. Ähnliches kann für „mops“ und „carlin“ gesagt werden, oder für „gazetta“ und „zeitung“.

Guy und Tonkin (vgl. [Guy2006]) untersuchten ca. 3000 Tags aus del.icio.us und Flickr. Sie prüften unter anderem die Sprache der einzelnen Tags mit *aspell*[13], einem Unix-Werkzeug zur Rechtschreibprüfung und textbasierter Spracherkennung.

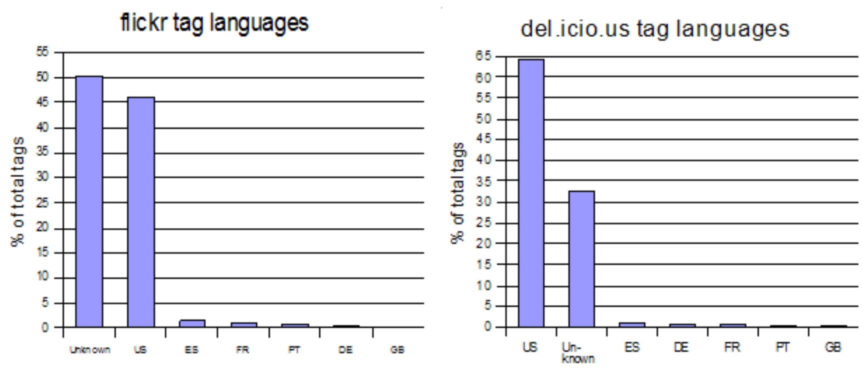


Abbildung 2-1: Erkannte Sprachen aus 3000 Tags von Flickr und del.icio.us

Wie in Abbildung 2-1 zu sehen ist, ist englisch mit Abstand die am häufigsten erkannte Sprache: 45% der Flickr-Tags und 65% der del.icio.us-Tags identifiziert aspell als englisch (American English oder British English werden zusammengefasst betrachtet, auch wenn es teils erhebliche Unterschiede in der Schreibweise vieler Wörter gibt. Wo es in British English beispielsweise „colour“ und „aluminium“ heißt, wäre in American English „color“ und „aluminum“ korrekt[14]. Die zweit häufigste Sprache ist Spanisch, die jedoch in beiden Systemen mit unter 2% kaum ins Gewicht fällt. Es folgen Französisch bzw. Deutsch und Portugiesisch, wobei das Vorkommen dieser Sprachen verschwindend klein ist.

Natürlich lässt sich daraus nicht schließen, dass es keine Sprachbarriere gäbe. Aus unserem deutschsprachigen Kontext heraus haben wir eine andere Sicht auf dieses Problem und empfinden es eher als störend als es Menschen in einem englischsprachigen Land tun würden. Guy und Tonkin bemerken dazu:

> *Though the bulk of tags found in our samples from flickr and del.icio.us combined are valid English language dictionary words using US or British spelling, tags from other countries are represented and may be in various foreign languages.*

13 Siehe http://aspell.sourceforge.net/man-html/Supported.html. Zugegriffen am 04.07.2006.
14 Siehe http://miless.uni-duisburg-essen.de/servlets/DocumentServlet?id=11636. Zugegriffen am 10.07.2006.

Sie nennen ein weiteres Beispiel, für das Problem der Sprachbarriere. Manche Worte kommen, in gleicher Schreibweise, in verschiedenen Sprachen vor. So ist „son“ nicht nur das englische Wort für „Sohn“, sondern auch ein adjektivisches Possessivpronomen in der französischen Sprache. Das im italienischen für Butter stehende Wort „burro“ meint auf spanisch einen Esel.

iii Rechtschreibfehler

Guy und Tonkin haben bei einer automatischen Überprüfung von insgesamt ca. 3000 Tags aus del.icio.us und Flickr herausgefunden, dass 40% der Flickr-Tags und 28% der del.icio.us-Tags Rechtschreibfehler enthielten, in einer Sprache, die nicht durch die von ihnen benutzte Software unterstützt wurde, geschrieben wurden oder aus mittels Symbolen zusammengesetzten Wörtern bestanden.

Ein Teil dieser „Rechtschreibfehler“ wird nach ihrer Beobachtung durch fehlerhaftes oder ungenügendes Character Encoding verursacht. Gerade bei Sprachen, deren Zeichensatz über den der lateinischen Zeichen hinausgeht, kommt es nach Darstellung von Guy und Tonkin zu Problemen. Angeführt werden zum Beispiel die deutsche Sprache mit ihren Umlauten, aber auch chinesisch, japanisch, russisch und tschechisch. Dieser Teil des Problems ließe sich durch eine konsequentere Nutzung von Unicode – einem Zeichensatz, der die Zeichen (fast) jeder lebenden Sprache abbildet - beheben oder zumindest besser beherrschbar machen.

Leider ist die Untersuchung an dieser Stelle ungenau. Aus den vorliegenden Zahlen und mittels der genutzten Methode lässt sich keine exakte Aussage über Rechtschreibfehler machen. Immerhin unterstützt aspell aber eine Vielzahl an Sprachen, so dass der Anteil der deswegen nicht erkannten Worte relativ klein sein dürfte. Guy und Tonkin untersuchten außerdem die zusammengesetzten Begriffe und nennen für diese – in den Daten von del.icio.us – einen relativen Anteil von 10%. Damit sinkt der Anteil der vermeintlichen Rechtschreibfehler auf 18% für die del.icio.us-Daten.

Ein Teil der als fehlerhaft erkannten Tags bei Flickr könnten sogenannte Geotags sein. Geotagging[15] gibt, mittels besonders kodierter Angaben zu Längen- und Breitengrad (bspw. „geo:lat=47.559967“ und „geo:lon=7.591526“) den exakten Ort, an dem eine Fotografie gemacht wurde, wieder. Desweiteren werden Fotografien bei Flickr oft mit Angaben zur Kamera[16], zum Beispiel der Modellnummer („d50“ [140.369], „20d“ [106.784]) und -bezeichnung („eos“ [102.777], „coolpix“ [25.001], ..) und Herstellernamen („canon“ [502.507], „kodak“ [90.165], „nikon“ [337.472], ...) getaggt.

Der im Vergleich zu del.icio.us wesentlich höhere Anteil an nicht erkannten Tags (40% gegenüber 28%) könnte also auch aus diesen quasi nicht erkennbaren Tags herrühren.

iv Sonderzeichen in Tags

Ein weiteres von Guy und Tonkin erwähntes Problem sind Symbole wie # oder @, die benutzt werden, um einzelne Tags in der alphabetischen Liste direkt zu Beginn auftauchen zu lassen. Auch solche Tags sind in die Summe der nicht erkannten Tags einzurechnen. Beispielhaft seien hier drei Linksammlungen erwähnt, die der del.icio.us-Nutzer „brownorama“[17], „jamesmelzer“[18] und „erikhaugo“[19]. Beide nutzen bestimmte Sonderzeichen zur Markierung von Themen, Objektarten, Autoren, Quellen und Aktivitäten, die sie mit den Ressourcen ihrer Sammlung verbinden. Beide bewerten die Ressourcen auch durch Sonderzeichen, nämlich Sternen (*). In Anlehnung an die Tag-Typen nach Golder und Hubermann (siehe Seite 17ff) hier eine vergleichende Übersicht über Tag-Typen und wie sie mit Sonderzeichen markiert werden.

15 Siehe http://en.wikipedia.org/w/index.php?title=GeoTagging&oldid=62847755. Zugegriffen am 12.06.2006.
16 Siehe http://flagrantdisregard.com/flickr/topcameras.php – Die Angaben in eckigen Klammern entspricht der Anzahl der mit den jeweiligen Tags getaggten Bildern bei Flickr am 13.07.2006.
17 Siehe http://del.icio.us/brownorama/ - Zugegriffen am 13.07.2006.
18 Siehe http://del.icio.us/jamesmelzer – Zugegriffen am 13.07.2006.
19 Siehe http://del.icio.us/erikhaugo – Zugegriffen am 13.07.2006.

Identifying what it is	Tags die Angaben zum Typ einer Ressource machen, wird das Gleichheitszeichen vorangestellt: „**=excel**“ oder „**=review**“.
Identifying who it owns	Angabe der Quelle mit „**in:Quelle**“ oder „**@Quelle**“. Parallel dazu werden die Stellen, von denen aus man auf die Quelle stieß, mit „**via:Hinweisequelle**“ getaggt.
Identifying qualities or characteristics	Bewertung des Inhalts mit Sternen nach einem Punkteschema; je mehr Sterne (bis zu fünf werden i. d. R. vergeben), desto besser oder interessanter ist die Ressource. Zum Beispiel **„*****“** für eine sehr gute oder interessante Ressource. Adjektiven wird ein Tilde vorangestellt, zum Beispiel **„~funny“** oder **„~cool“**.
Task organizing	Solchen Tags wird ein Ausrufezeichen vorangestellt, zum Beispiel **„!toread“** oder **„!tobuy“**.

Tabelle 1: Nutzung von Sonderzeichen zur Markierung bestimmter Tag-Typen.

v Zusammengesetzte Begriffe

Guy und Tonkin identifizieren ein weiteres Problem als das größte mit einem Vorkommen von über 10% von allen untersuchten del.icio.us-Tags. del.icio.us Benutzer verwenden oft zusammengesetzte Worte als Tags. Diese werden aber, aufgrund einer fehlenden Konvention, mit verschieden Trennzeichen geschrieben.

Zumeist wird zwar der einfache Bindestrich genutzt, andere Zeichen tauchen aber ebenso in signifikantem Anteil auf. Dem Schrägstrich (Slash) kommt überdies eine weitere Bedeutung zu. Er wird benutzt, um einen (nach Guy und Tonkin) „pathway“ darzustellen.

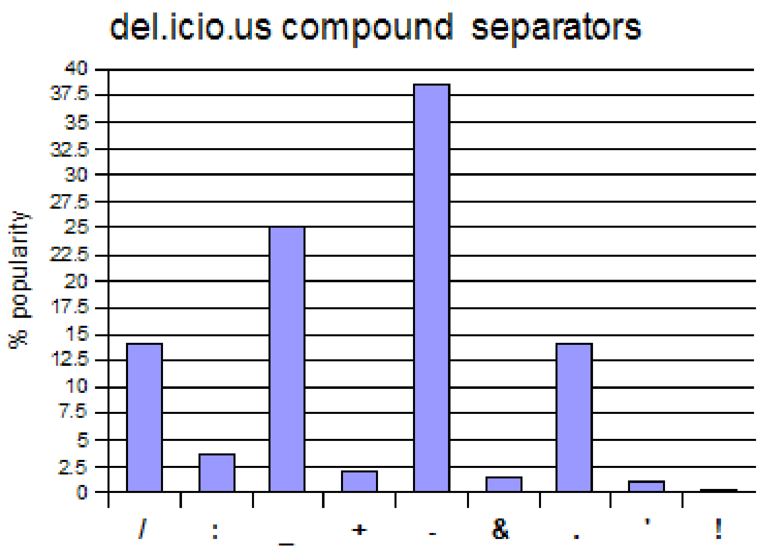


Abbildung 2-2: Relativer Anteil verschiedener Trennzeichen bei zusammen gesetzten Wörtern



Abbildung 2-3: Der Slash als Trennzeichen bei del.icio.us.

Beispielsweise „devel/c++“, „devel/php“ oder „devel/java“. Auch wenn del.icio.us diese „pathways“ nicht als solche abbildet – das heißt, del.icio.us kennt das Konzept einer Baumstruktur oder Hierarchie nicht; Tags werden schlicht alphabetisch geordnet – erscheinen sie in der alphabetischen Liste der eigenen Tags geordnet in einer „Baumstruktur“:



Abbildung 2-4: Mehrere Ebenen einer Pseudo-Baumstruktur mittels dem Slash als Trennzeichen.

Diese „Baumstruktur“ lässt auch mehrere Ebenen zu. del.icio.us ermöglicht es aber nicht, da es keine Hierarchie kennt, diese „pathways“ zu trennen. Eine Navigation von „GM“ zu „Fehler“ zu „test“ ist nicht möglich – für das System ist „GM/Fehler/test“ *ein* Tag.

Neben der Benutzung von Trennzeichen zur Zusammenführung mehrerer Begriffe zu einem Tag wird auch der von den Wikis entliehene Ansatz des CamelCase[20] benutzt. Wikis sind sehr einfache Web Content Management Systeme, die in der Grundform jedem die Veränderung des Inhalts einer Seite erlauben. Wikis nutzen CamelCase zur Beschreibung von Seitennamen. Erkennt die Wiki Engine[21] einen Ausdruck in CamelCase, verlinkt es automa-

20 Siehe http://c2.com/cgi/wiki?CamelCase – Der Ausdruck CamelCase rührt aus der Praxis her, in Programmcode statt Unterstrichen gar keine Trennzeichen zu verwenden: *READ_WRITE_STREAM* wird zu *ReadWriteStream*. Zugegriffen am 11.07.2006.

21 Siehe http://c2.com/cgi/wiki?WikiEngine – Eine Wiki Engine ist eine Software, die nach den Wiki-Prinzipien (http://c2.com/cgi/wiki?WikiPrinciples) funktioniert. Zugegriffen am 11.07.2006.

tisch auf diese Seite. Begriffe in CamelCase sind also sowohl Seitennamen bzw. -titel als auch Teil der URI der Seite.

Zu den bereits erwähnten nicht erkennbaren Tags in Bezug auf Guys und Tonkins Untersuchung zu „falschen“ Tags müssen wir auch diejenigen Tags zählen, die in CamelCase geschrieben sind, es sei denn, der so zusammengesetzte Begriff löst sich zu einem Wort einer bekannten Sprache auf.

vi Subjektivität der Information

Informationen an sich sind nicht subjektiv, wohl aber der Informationsbedarf des Nutzers. Es kann also durchaus vorkommen, dass mehrere Nutzer die selbe Ressource aus unterschiedlichen Interessen oder Kenntnissen oder Aufgabenstellungen (vgl. [Fugmann1999]) heraus taggen. Aufgrund der unterschiedlichen Motivationen, Wissensstände und äußeren Umstände ist es sehr unwahrscheinlich, dass die Nutzer die exakt gleichen Tags wählen. Wahrscheinlich werden sie im Kern gleiche oder ähnliche Tags wählen, aber um diese Kern-Tags herum solche, die es ihnen selbst leichter machen, später den Inhalt wieder einzuschätzen.

Golder und Huberman weisen ebenfalls auf dieses Problem hin und beschreiben drei Klassen tatsächlich persönlicher Tags (Identifying qualities or characteristics, Self reference, Task organizing [siehe Kapitel Identifying qualities or characteristics, Self reference und Task organizing, S. 18ff]), die nur für denjenigen Benutzer Sinn ergeben, der sie selbst eingegeben hat. Im Sinne eines kollaborativen Ansatzes (siehe Kollaborativer Ansatz, S. 38ff) sind solche Tags nutzlos und unterstützen weder das Browsing noch die Suche.

Mejias (siehe Kapitel Vorschläge zur Verbesserung des Tagging-Prozesses, S. 37f) (vgl. Mejias2005]) zitiert an dieser Stelle Kress (vgl. [Kress2003]):

> *[...] words in combination are not much more than rough outlines waiting for us the readers to colour them in. What the written text provides is words in clear order. Each word asks to be filled with meaning, a*

> *meaning that comes from our past experience of that word in our social lives.*

Damit gemeint ist, dass ein grammatikalisch und syntaktisch korrekter Satz dennoch von jedem Leser anderes interpretiert wird. Diese Interpretation basiert auf den Erfahrungen des Lesers und seinem Wissen. Ein gutes Beispiel dafür sind Gedichte oder ironische Bemerkungen. Mejias führt dazu weiter aus:

> *While Kress is referring to written texts with a "clear order" (such as sentences and paragraphs), an unordered group of tags used to describe a resource holds the same potential to be filled with personal and social meaning. Clearly, these groups of words are not devoid of meaning as texts. To their authors, they represent textual associations that will make it easier to find the resource again in the future (one can assume, for example, that the tag to_read is used to classify resources that the user intends to read later). Each group of tags is the user's framing of the resource according to a personal scheme. At the same time, however, the fact that certain keywords (such as folksonomy, social, del.icio.us, ontology, etc.) are repeated across groups suggests that some of these words have a socially shared meaning apart from their personal meaning. This socially shared meaning is what would allow someone browsing through those tags to find that same resource, or similar ones.*

Tags sind also nicht nur freie Schlagworte, die den Inhalt einer Ressource beschreiben, sondern auch persönliche Referenzen für einzelne Nutzer, die ihnen helfen, bekannte Inhalte zu einem zukünftigen Zeitpunkt wieder zu finden. Die zentrale Aussage von Mejias zu diesem Punkt ist, dass jede Gruppe von Tags (diejenigen, die von einem Benutzer für eine Ressource genutzt wurden) eine Repräsentation dieser Ressource aus Sicht dieses einen Benutzers ist. Sinah (vgl. [Sinah2005]) schreibt dazu:

Cognitive psychologists have explored this phenomenon by asking [by asking] people to list semantic associations with an object, and mapping the type and frequency of associations. Another method is to use implicit memory measures to probe what concepts have been activated. With the advent of fMRI[[22]], it is possible to correlate such concept activation to changes in blood flow to difference parts of the brain. The details of this are not relevant for the present discussion, what's relevant is that there is broad agreement about such conceptual activation in cognitive psychology.

Er beschreibt den Taggingprozess mit dieser Abbildung:

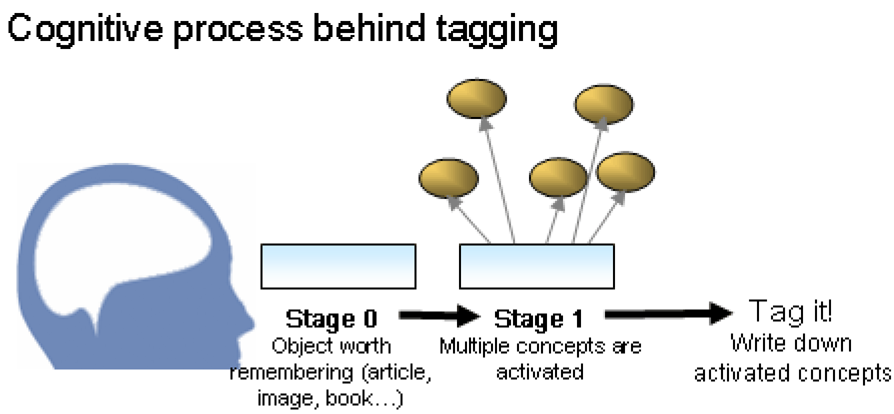


Abbildung 2-5: Ablauf des Taggingprozess nach Sinah.

c Motivation

Marlow et al. (vgl. [Marlow2005]) unterteilen die grundlegenden Motivationen von Nutzern, Taggingsysteme zu nutzen, in zwei Klassen: *organizational* und *social*. Sie verstehen darunter zum einen die Motivation, eine strukturierte Ablage für Ressourcen zu schaffen. Solche Nutzer entwickelten einen persönlichen Standard, angelehnt an die im System vorherrschenden Nutzungsverhalten:

In einem Video[23] zeigt Udell, wie er in del.icio.us eine Ressource mit dem Tag „digitalidenity“ versieht. Er prüft seine übrigen mit diesem Tag versehenen Ressourcen und stellt fest,

22 **Functional magnetic resonance imaging – funktionelle Kernspintomographie. Siehe** http://en.wikipedia.org/w/index.php?title=Functional_magnetic_resonance_imaging&oldid=68390787 – Zugegriffen am 11.08.2006.
23 Siehe http://weblog.infoworld.com/udell/gems/delicious.html – Zugegriffen am 17.07.2006

das zwei dieser Ressourcen von anderen Nutzern nicht mit „digitalidenity“, sondern „cert“ für *certificate* getaggt wurden. Daraufhin löscht er für diese beiden Ressourcen das Tag „digitalidentity“ und ersetzt es durch „digitalcertificate“. Er stellt auch fest, dass er in der Vergangenheit statt „digitalidenity“ das Tag „identity“ verwendet hat und ersetzt daraufhin alle Vorkommen von „digitalidentity“ durch „identity“.

Zum anderen gibt es Nutzer, die vorwiegend sozial motiviert sind. Sie nutzen Taggingsysteme, um ihre Meinungen über Ressourcen (z. B. „funny“, „boring“) wiederzugeben und Besonderheiten (z. B. „violence“, „adult“) hervorzuheben.

Beide Ansätze unterscheiden sich in dem Zielpublikum sowie in der Erwartung der späteren Nutzung durch den Nutzer selbst. Marlow et al. beschreiben sechs verschiedene Motivationen:

1. Future retrieval
2. Contribution and sharing
3. Attract Attention
4. Play and competition
5. Self presentation
6. Opinion expression

i Future retrieval

Das Taggen von Ressourcen zur späteren Nutzung, entweder einzelner Ressourcen oder einer Menge mehrerer Ressourcen (zum Beispiel in Vorbereitung einer Diplomarbeit). Tags der Klasse Task organizing (siehe S. 18), die konkrete Aufgaben beschreiben, wie „toread“ und ähnliche, wurden aus dieser Motivation heraus vergeben.

ii Contribution and sharing

Damit gemeint ist die Motivation, Ressourcen für andere bereit zu stellen, um sie auf etwas aufmerksam zu machen oder sie auf etwas hinzuweisen. Beispielsweise das Taggen von Reiseinformationen für den Partner oder Fotografien einer Veranstaltung: Der Web Montag[24], der seit Anfang des Jahres in mehreren deutschen Städten stattfindet, weist auf seiner Website darauf hin, wie man in den verschiedenen Taggingsystemen taggen (i. d. R. mit dem tag „webmontag", oder auch „webmontag20060814" [Dieses Tag steht für den kommenden Webmontag am 14.08.2006.]) soll, damit andere zusammengehörige Inhalte finden können. Im Grunde geht es um eine Vereinbarung innerhalb einer Gruppe von Menschen, für bestimmte Inhalte das gleiche Tag zu nutzen, um das Auffinden zusammengehöriger Informationen zu erleichtern.

iii Attract Attention

Die Motivation, Aufmerksamkeit zu erzeugen, spiegelt sich in Taggingsystemen darin wieder, dass so motivierte Nutzer vor allem solche Tags vergeben, die einen hohen Grad an Aufmerksamkeit genießen. Tag Clouds oder ähnliche Listen der meist genutzten Tags geben diesen Nutzern Hilfestellung. Auch kann die Motivation darin liegen, ein Tag in diesen Listen und Darstellungen zu „pushen" oder ein gänzlich neues einzuführen.

iv Play and competition

Tags nach einem Regelwerk zu vergeben, wie im ESP Game[25], wo der Nutzer ein Bild so taggen sollen, wie er meint, dass andere Nutzer es auch tun, oder wie zum Beispiel bei Odeo[26], als Odeo-Nutzer einen Versuch starteten, Tags so zu vergeben, dass sich in der Tag Cloud Sätze bildeten.

24 Siehe http://www.webmontag.de/ - Zugegriffen am 11.08.2006.
25 Siehe http://www.espgame.org/ - Zugriffen am 17.07.2006.
26 Siehe http://www.odeo.com/ - Zugegriffen am 17.07.2006.

v Self presentation

Tags, die aus dieser Motivation heraus vergeben werden, beschreiben den Bezug des Nutzers zur getaggten Ressource. Zum Beispiel das Tag „seen live“ auf last.fm[27] - einer Plattform zum Taggen vom Musikstücken und Künstlern. Es beschreibt die Tatsache, dass der Nutzer dieses Musikstück oder diesen Künstler in einem Konzert erlebt hat. Tags der Klasse Self reference (siehe S. 18), die die Beziehung zwischen Nutzer und Ressource beschreiben, wurden aus dieser Motivation heraus vergeben.

vi Opinion expression

Findet ein Nutzer eine Ressource besonders nützlich (oder unnütz), lustig, informativ, spannend u. ä., möchte er seine Meinung mit anderen Nutzern teilen und benutzt dafür Tags der Klasse Identifying qualities or characteristics (siehe S. 18) wie zum Beispiel „funny“, „boring“ oder „well-written“.

d Ausprägungen von Taggingsystemen

Betrachtet man die verschiedenen Taggingsysteme, fällt auf, dass sie zwar alle auf Tagging basieren, sich in der Implementierung einzelner Funktionen jedoch stark unterscheiden. Marlow et al. haben aus diesen Beobachtungen eine Taxonomie von Taggingsystemen erarbeitet, die im folgenden dargestellt wird.

Sie unterscheiden sieben Dimensionen:

1. Tagging Rights

2. Tagging Support

27 Siehe http://last.fm/ - Zugegriffen am 17.07.2006

3. Aggregation
4. Type of object
5. Source of material
6. Resource connectivity
7. Social connectivity

i Tagging Rights

Diese Dimension macht eine Aussage darüber, welcher Nutzer welche Ressourcen taggen und die Tags verwalten darf. Sie spannt sich auf von restriktiven Systemen, in denen nur derjenige Nutzer eine Ressource taggen darf, der sie zuerst dem System bekannt macht („self-tagging" nach Marlow et al.), bis zu offenen („free-for-all" nach Marlow et al.) Systemen, die jedem Benutzer erlauben, jede Ressource zu taggen. Die Unterschiede in den einzelnen Implementierungen sind fein granuliert und folgen dem gesamten Bogen. Ein ähnlich weiter Bogen überdacht die Ausprägungen der Antwort auf die Frage, wer ein Tag löschen darf. Die Positionierung eines Taggingsystems innerhalb dieser Dimension hat direkte Auswirkungen auf die Tags, die in ihm benutzt werden. „free-for-all"-Systeme ordnen einer Ressource viele Tag-Gruppen zu, wobei die Tags selbst sehr unterschiedlich sein können. „self-tagging"-Systeme hingegen ordnen jeder Ressource nur eine einzige Tag-Gruppe zu. So kommt es in solchen Systemen vor allem darauf an, wer und aus welcher Motivation heraus derjenige eine Ressource taggt. Der Fotograf selbst würde einer Fotografie andere Tags mitgeben als eine beliebige dritte Person es tun würde.

Dimension	Main categories	Summary of potential implications
Tagging Rights	self-tagging, permission-based, free-for-all	Nature and type of resultant tags; role of tags in system
Tagging Support	blind, suggestive, viewable	Convergence on folksonomy or overweighting of tags
Aggregation model	bag, set	Availability of aggregated statistics
Object type	textual, non-textual	Nature and type of resultant tags
Source of material	user-contributed, system, global	Different incentives, nature and type of resultant tags
Resource connectivity	links, groups, none	Convergence of similar tags for linked resources
Social connectivity	links, groups, none	Convergence on localized folksonomy

Tabelle 2: Dimensionen des Designs von Taggingsystemen nach Marlow et al.

Hammond et al.[24] ordnen Taggingsysteme nach der Frage, wer den Inhalt erstellt hat sowie wer den Inhalt taggen darf. Sie benutzen dazu das Diagramm aus Abbildung 2-6, S. 33.

Nach Hammond et al. bewegen sich Nutzungsszenarien aus dem unteren linken Quadranten, in dem eigene Inhalte für eigene Zwecke getaggt werden (selfish) in den oberen rechten Quadranten (altruistic), während sich die meisten Taggingsysteme im unteren rechten Quadranten wiederfinden, in dem Nutzer fremde Inhalte für eigene Zwecke taggen. Der obere linke Quadrant nimmt HTML Meta Tags sowie Technorati auf, beides Systeme um eigene Inhalte für andere zu erschließen. Wikipedia benutzt ein Kategorienschema zur Inhaltserschließung, keine Tags. Warum Hammond et al. es dennoch erwähnen, bleibt offen.

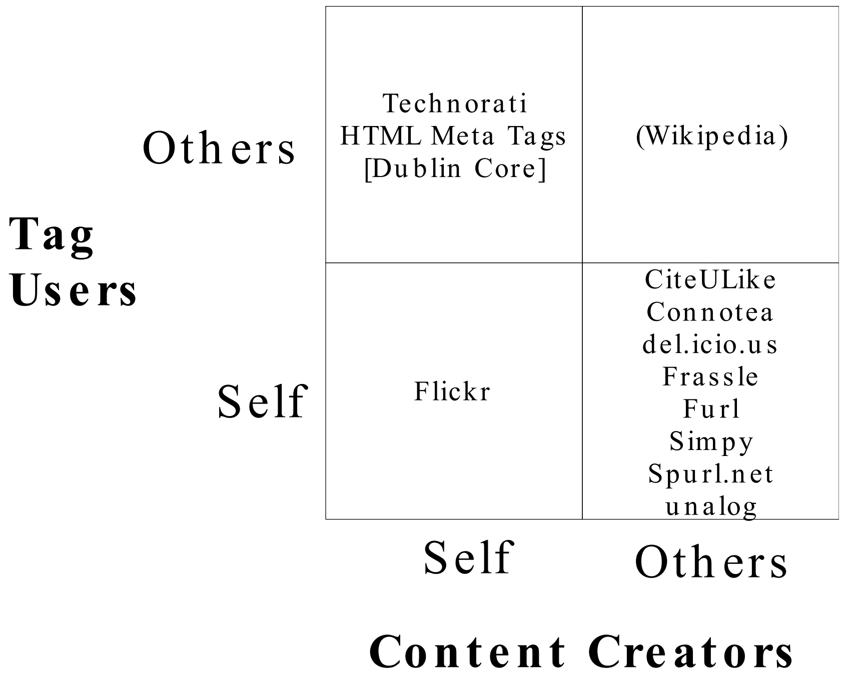


Abbildung 2-6: Einordnung der Taggingsysteme nach Hammond et al.

Die Systeme im oberen rechten Quadranten, Technorati[28], HTML Meta Tags und auch Metadaten nach Dublin Core sind im eigentlichen Sinne keine Tags, da sie nicht in einer Folksonomy zusammengefasst werden können. Technorati-Tags bilden hier eine Ausnahme, weil Technorati als zentrale Stelle diese Tags durchaus zusammenfasst. Jedoch besitzen sie nur in Weblogs Relevanz und sind daher als spezialisierte Version von HTML Meta Tags oder Angaben nach Dublin Core zu sehen.

Thomas Vander Wal (vgl. [VanderWal2005]) beschreibt „broad“ und „narrow“ Folksonomies. Auch wenn Marlow et al. Vander Wals Beschreibung nicht erwähnen oder referenzieren, so gleicht ihre Definition von „free-for-all“ und „self-tagging“ doch sehr Vander Wals „broad“- und „narrow“- Ansätzen. Vander Wal fügt seiner Begriffserklärung zwei Darstellungen (Abbildungen 2-7 und 2-8) bei, die das Konzept einer „broad“ bzw. „free-for-all“ und „narrow“ bzw. „self-tagging“ Folksonomy verdeutlichen.

---

28 Technorati ist eine Weblog-Aggregator, der aus den publizierten Feeds von Weblogs eine Suche bereitstellt. Technorati-Tags sind eine für diese Website und ihre Dienste spezifische Anwendung.

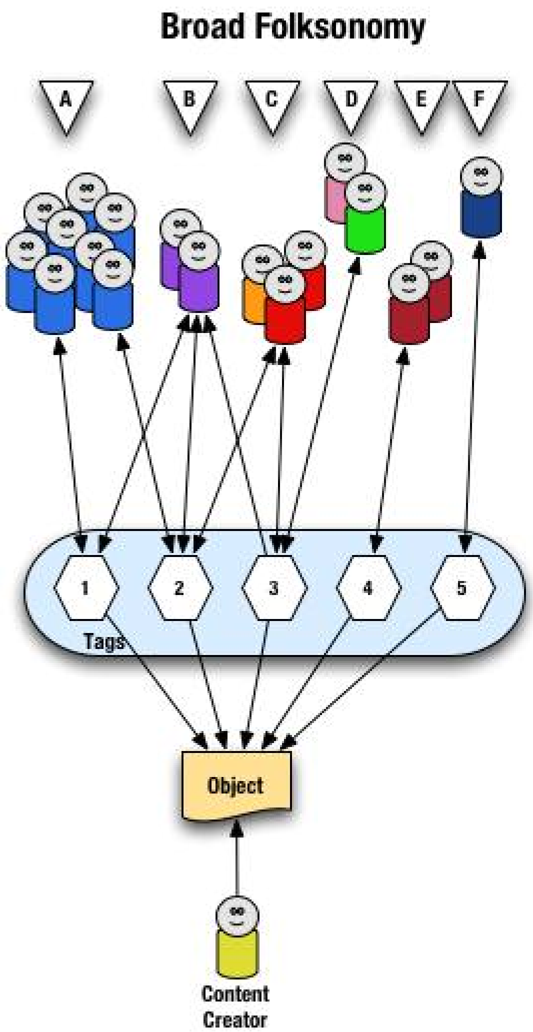


Abbildung 2-7: Darstellung des Modells einer "broad folksonomy" nach Vander Wal.

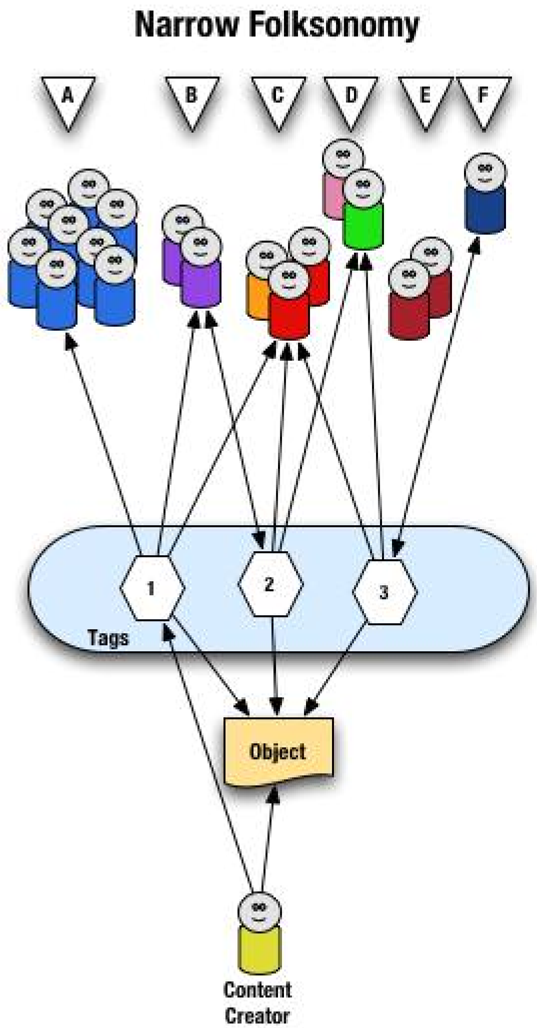


Abbildung 2-8: Darstellung des Modells einer "narrow folksonomy" nach Vander Wal.

ii Tagging Support

Diese Dimension hat nach Marlow et al. drei Ausprägungen: „blind tagging“, „viewable tagging“ und „suggestive tagging“.

Hat der Nutzer, während er eine Ressource taggt, keinen Zugriff auf Tags, die von anderen Benutzern für die gleiche Ressource vergeben wurden, wird dies „blind tagging“ genannt. Als „viewable tagging“ bezeichnet man die Situation, wenn der taggende Nutzer die Tags anderer Nutzer sehen kann. „suggestive tagging“ schließlich bezeichnet ein System, das

dem Nutzer auf Basis seiner früher eingegebenen oder von anderen für diese Ressourcen vergebenen Tags Vorschläge macht. Vorgeschlagene Tags können auch aus automatisch extrahierten Metadaten stammen oder können automatisch erkannte Synonyme anderer Vorschläge sein.

Auch Golder und Huberman gehen auf dieses Thema ein. Sie sagen:

> *A user may use the suggested popular tags as a way of looking to others to see what the "right" thing to do is.*

Golder und Huberman erklären diese Überlegung mit einem Verweis auf das Prinzip des „social proof", einer Hypothese von Cialdini (vgl. [Cialdini2001]): Menschen tendieren dazu, die Handlungen und Handlungsweisen der Menschen in ihrer direkten Umgebung nachzuahmen.

Daraus schließen Marlow et al., dass suggestive Systeme helfen könnten, konvergentere Folksonomies zu bilden, da sie Nutzern dazu raten, bereits von anderen genutzte Tags zu vergeben.

iii Aggregation

Systeme unterscheiden sich auch in der Art und Weise, welche Tags sie je Ressource speichern. Wo manche Systeme eine beliebige Zahl von Tag-Gruppen zulassen – nach Marlow et al. ein „bag-model" – erlauben andere nur eine einzige Tag-Gruppe – ein „set-model". Vander Wal hat diesen Ansatz ebenfalls in seine beiden Modelle einfließen lassen (siehe Abbildungen 2-7 und 2-8).

iv Type of object

Nachdem zunächst Webseiten getaggt wurden, kamen schnell andere Inhalte hinzu. Bilder, Videos und andere Ressourcen, die online repräsentiert (direkt oder indirekt, zum Beispiel in Katalogen von Museen (siehe Abschnitt steve.museum, S. 12f) sind. Auch Objekte

der physikalischen Welt werden getaggt[29], genauso wie Veranstaltungen aller Art[30]. Auch innerhalb einer Klasse, zum Beispiel Webseiten, werden mittlerweile Unterschiede gemacht. Wo del.icio.us jeden per URL erreichbaren Inhalt zulässt, wird CiteULike dazu benutzt, ausschließlich wissenschaftliche Veröffentlichen in sogenannten „libraries“ zu sammeln.

Auch wenn es bislang keine Untersuchungen zu diesem Thema gibt, vermuten Marlow et al., dass der Typ der Ressource eine Auswirkung auf die Tags hat, die ihm von den Nutzern zugeordnet werden.

v Source of material

Dieser Aspekt unterscheidet die Quellen einer Ressource nach Hersteller oder Besitzer, dem System, und einer grundsätzlichen Offenheit. Systeme wie Flickr erlauben nur den Hersteller oder Besitzer als Quelle, das ESP Game stellt von sich aus Inhalte bereit und del.icio.us erlaubt es, dem System beliebige Webinhalte hinzuzufügen.

vi Resource connectivity

Konnektivität drücken Marlow et al. in den Begriffen „linked“, „grouped“ und „none“ aus. Beispielsweise sind zwei Webseiten „linked“, wenn sie auf einander verweisen. Innerhalb von Flickr kann man Fotografien Gruppen zuordnen, was „grouped“ entspräche. „none“ bedeutet, das es keine Möglichkeit gibt, Ressourcen zueinander in Verbindung zu bringen.

Marlow et al. äußern die Vermutung, dass die Verbindung von Ressourcen zu einer Konvergenz der verwendeten Tags führen könnte, vor allem in „suggestive tagging“- und „viewable tagging“-Systemen, die nach dem „bag-model“ arbeiten, also mehrere Tag-Gruppen pro Ressource zulassen.

29 Siehe http://www.semapedia.org/ - Zugegriffen am 11.08.2006.
30 Siehe http://upcoming.org/ - Zugegriffen am 11.08.2006.

vii Social connectivity

Die gleichen Begriffe betreffen auch die Verbindung der Nutzer untereinander. Hier identifizieren Marlow et al. Unterdimensionen, so zum Beispiel den Typ einer Verbindung zwischen zwei Nutzern (Freund oder Kontakt[31], reine Beobachtung[32], ...), ob die Verbindungen gerichtet sind und ob sie symmetrisch sind bzw. sein müssen.

Die Auswirkungen der Dimension des sozialen Systems bzw. des gegenseitigen Bezugs der Nutzer eines Taggingsystems könnten die Bildung lokalisierter Folksonomies basierend auf der sozialen Struktur des Gesamtsystems sein.

e Vorschläge zur Verbesserung des Tagging-Prozesses

Es liegen einige Vorschläge zur Verbesserung von Tags und des Taggingprozesses selbst vor. Mejias zum Beispiel hält eine ganze Reihe von Maßnahmen für nützlich.

Zunächst schlägt er vor, Tags, die als Kategorien dienen, im Plural anzulegen. Weiterhin sollte man grundsätzlich Kleinschreibung verwenden und Großschreibung nur da, wo es unbedingt notwendig ist, zum Beispiel bei Eigennamen (von Personen und Institutionen) oder Abkürzungen. Benutzer von Taggingsystemen (von Mejias DCS – Distributed Classification Systems genannt) sollten nicht nur spezielle Tags nutzen, sondern auch allgemeine Begriffe verwenden, um anderen Benutzern das Auffinden einer Ressource zu erleichtern. Er rät auch dazu, persönliche Tags (toread, print_this, funny, clever, ...) nicht ausschließlich zu nutzen, sondern immer auch für andere Nutzer sinnvolle Tags zu verwenden. Mehrwortige Tags wie „open source“ sollten nicht in ihre Bestandteile gespalten werden, um die Wiederauffindbarkeit besser zu gewährleisten. Technisch besteht hier noch das Problem, dass nicht alle Taggingsysteme solche Zusammenführungen erlauben und es noch keinen Standard gibt, solche Zusammenführungen abzubilden. Wie im Abschnitt Zusammengesetzte Begriffe, S. 23f, gezeigt, hat sich noch kein Standard etabliert, wie zusammengesetzte Begriffe getrennt werden sollten. Mejias schlägt auch vor, synonyme Begriffe für einzelne Tags zu benutzen. Gerade

31 Möglich bei Flickr.
32 Das del.icio.us-„network“ basiert auf dieser Idee. Siehe Abschnitt Navigation in del.icio.us, S. 67ff.

bei Begriffen, von denen klar ist, das sie viele akzeptierte Synonyme haben sei dies sinnvoll. Letztlich ermahnt Mejias Benutzer von Taggingsystemen, ihr System genau zu beobachten und sich an bestehende Regeln (die meist nicht ausformuliert sind sondern oft „ungeschriebene Gesetze“ sind) zu halten und schließlich die Möglichkeiten zur Verbesserung der eigenen (und fremder, wenn vom System erlaubt) Tags zu nutzen und so zur Verbesserung des gesamten Systems für alle Nutzer beizutragen.

Melzer (vgl. [Melzer2005]) schlägt vor, bestimmte Tag-Typen mit Sonderzeichen zu versehen. So könne man zum Beispiel persönliche Tags wie „toread“ mit einem Ausrufezeichen versehenen (für eine zusammenfassende Übersicht siehe Sonderzeichen in Tags, S. 22f). Er rät außerdem, sowohl allgemeine als auch spezialisierte Tags zu benutzen, um die so getaggten Ressourcen später besser einordnen zu können.

Wie im Abschnitt Tagging Support, S. 34, beschrieben, gibt es unterschiedliche Ansätze, den Nutzer bei dem Tagging-Prozess zu unterstützen. Melzer (vgl. [Melzer2006]) schlägt dazu eine Erweiterung der Benutzeroberfläche vor, bezogen auf del.icio.us. Zum einen sollten dem Benutzer diejenigen Tags angezeigt werden, die er in der Vergangenheit bereit für Ressourcen der gleichen Quelle benutzt hat („Tags suggested based on tags I've used on this domain before“), zum anderen sollte del.icio.us eine Liste derjenigen Benutzer bereitstellen, die die gleiche Ressource bereits getaggt haben. In dieser Liste soll auch angezeigt werden, ob die Nutzer zu den Top 10, 50, 100 oder 1000 aller del.icio.us-Nutzer zählen. Melzer dazu:

> *Personally, I like to know how many other folks have bookmarked the item already. [...] Knowing who else tagged the item gives me a sense of whether this is a new meme or something that has already peaked. I also use this to find good taggers and early taggers and follow their links.*

f Kollaborativer Ansatz

Tagging ist fast immer als kollaborativer Ansatz implementiert. Das bedeutet zum einen, dass das System nach dem „bag-model“ arbeitet und zum anderen, dass jeder Nutzer jede Ressource taggen und so in seine Sammlung aufnehmen kann. Der Nutzen dieser Systeme

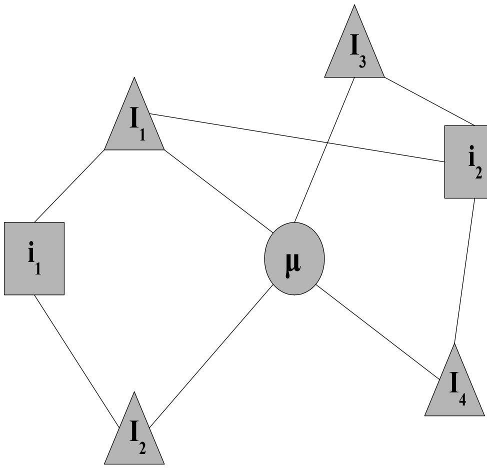
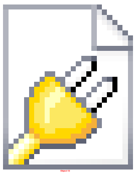


Abbildung 2-9: Graph eines dreiseitiges Netzwerks nach Lambiotte und Ausloos, das den Ansatz von Taggingsystemen zeigt.

steigt mit der Zahl der teilnehmenden Benutzer. Je nach Implementierung können Benutzer alle Ressourcen sehen, die das System kennt, gemeinsam mit allen Benutzern, die eine Ressource getaggt haben sowie die Tags selbst, die dabei verwendet wurden. In diesem Sinne ist Tagging zumeist ein dreiseitiges Netzwerk (vgl. [Lambiotte2005]):

> *The structure of collaborative websites can be viewed as a tripartite network. Namely, it is a network composed if three kind of nodes: i) the persons or users μ; ii) the items* i *that can be music groups or scientific articles; iii) the tags* I *that are used by the person* µ *to describe the item* i*. Depending on the systems under consideration, a person can use one or several tags on each item. The resulting network can be represented by a graph where edges run between the item* i *and the user* µ*, passing through the tag* I.

Darüberhinaus weisen sie den Verbindungen zwischen Benutzern und Ressourcen („items“) eine Wertigkeit in Abhängigkeit der Anzahl der benutzten Tags zu: . steht dabei für die Anzahl der Tags, mit der μ eine Ressourcen getaggt hat. So kann jeder Nutzer μ als Matrix von dargestellt werden:



Lambiotte und Ausloos schlagen auf dieser Grundlage Methoden vor, die aus Strukturen in Folksonomies Aussagen über Benutzergruppen (Benutzer, die ähnliche Inhalte ähnlich taggen) und zusammengehörige Ressourcen vor.

Mika [Mika20005] beschreibt eine Folksonomy als Hypergraph . V ist dabei die Vereinigungsmenge von A (der Menge der „Actors“, der Benutzer), C (der Menge der „Concepts“, der Tags) und I (der Menge der „objects“, der getaggten Ressourcen): . T ist dabei die Menge der Verbindungen zwischen Benutzern, Tags und Ressourcen: . E entspricht den Dreifachkanten zwischen allen Elementen von T: .

Aus dieser Betrachtungsweise ergeben sich nach Mika Ansätze zur Extraktion von Ontologien aus Folksonomies.

Marlow et al. treffen im Grunde die gleiche Aussage, wenn sie sie auch weniger mathematisch betrachten als Lambiotte und Ausloos. Sie fügen jedoch eine weitere Dimension hinzu, in dem sie die Möglichkeit offenlassen, dass es sowohl Verknüpfungen zwischen Ressourcen als auch zwischen Benutzern geben kann:

> *In this model, users assign tags to a specific resource; tags are represented as typed edges connecting users and resources. Resources may be also be connected to each other (e.g., as links between web pages) and users may be associated by a social network, or sets of affiliations (e.g., users that work for the same company).*

Nach diesem Modell sind Verknüpfungen in Form von Links zwischen Ressourcen ebenso zu beachten wie mögliche Verbindungen zwischen Benutzern.

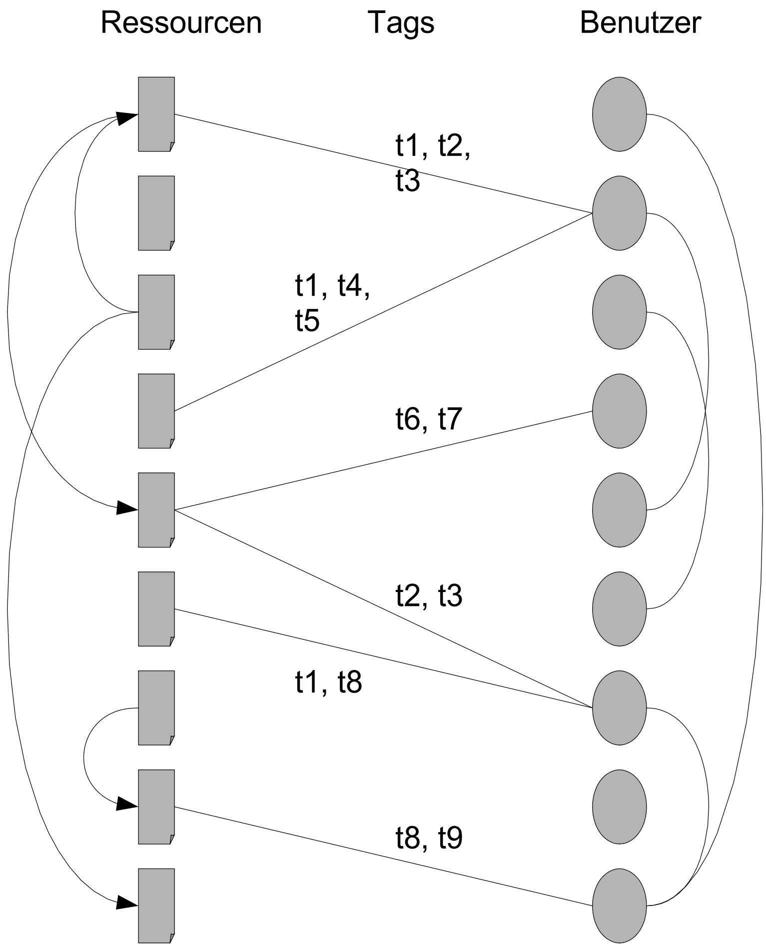


Abbildung 2-10: Tagging-Modell nach Marlow et al. Je eine "Tag-Gruppe" stellt eine Verbindung zwischen einer Ressource und einem Benutzer her. Darüberhinaus können sich durch Verlinkung innerhalb der Ressourcen (Hyperlinks) Verbindungen zwischen Ressourcen ergeben sowie durch Zugehörigkeit zu Organisationen Verbindungen zwischen Benutzern.

In Taggingsystemen ergeben sich also Beziehungen zwischen Benutzern und Ressourcen durch die von den Benutzern vergebenen Tags. Zu beachten ist, dass dies nur für Taggingsysteme gilt, die einen „bag-model“ Ansatz haben. Bei „set-model“ Ansätzen geht die Informat-ion, welcher Benutzer ein bestimmtes Tag benutzt hat, verloren.

CiteULike benutzt ein „bag-model“. Es verknüpft also einen Benutzer über eine Tag-Gruppe eindeutig mit einer Ressource. Als Tag-Gruppe wird die Menge von Tags bezeichnet, die ein Benutzer einer Ressource zuordnet:

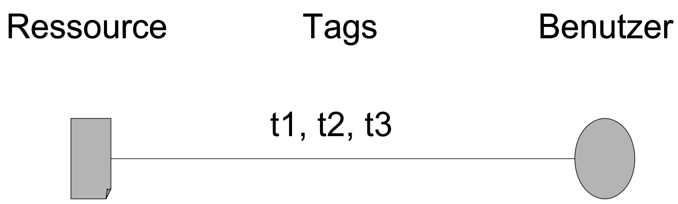


Abbildung 2-11: Eine Tag-Gruppe, bestehend aus den Tags t1, t2 und t3, verbindet eine Ressource und einen Benutzer.

Im Folgenden wird gezeigt, wie ein Werkzeug aussehen könnte, das anhand der Verbindungen zwischen Benutzern, Ressourcen und Tags Ressourcen finden kann, die ein Nutzer

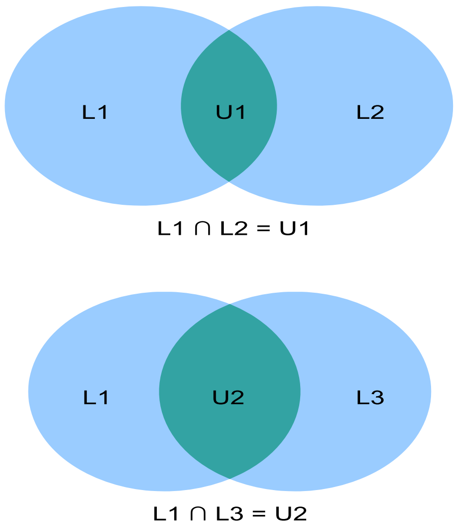


Abbildung 2-12: Schnittmengen dreier Bibliotheken L1, L2 und L3. Die Schnittmenge U1 umfasst weniger Elemente als U2. Damit sind die Elemente aus L3 \ L1 interessanter für mich als die L2 \ L1.

noch nicht kennt, aber potentiell für ihn interessant sein könnten.

i Collaborative Filtering auf Basis von Tagging

Die vier Hypothesen des im folgenden beschriebenen Ansatzes lauten:

1. „Teilt ein Benutzer viele Ressourcen mit mir, ist seine Bibliothek (L3) interessanter für mich als die eines anderen Benutzers, dessen Bibliothek (L2) eine kleinere Schnittmenge mit meiner (L1) hat (siehe Abbildung 2-12).“

2. „Ein Benutzer wird inhaltlich ähnliche Ressourcen immer gleich taggen oder zumindest ähnlich genug, dass man aufgrund der Tag-Gruppe der einzelnen Ressourcen zwischen sehr häufig und eher selten vergebenen Tags unterscheiden kann.“

3. „Nicht die gesamte Bibliothek L3 ist für mich interessant, sondern nur der Teil, den ich noch nicht kenne (L3 \ L1), und der mit ähnlichen Tags getaggt wurde wie die Ressourcen aus der Schnittmenge L1 ∩ L3.“

4. „Je mehr Bibliotheken eine bestimmte Ressource beinhalten, desto größer ist das Interesse an dieser Ressource.“

Um aus der Menge I = L3 \ L1 diejenigen Ressourcen zu identifizieren, die interessant sein könnten, wird folgendes Vorgehen vorgeschlagen.

1. Zunächst betrachtet man alle Elemente aus L1 und prüft für jede Ressource, in welchen anderen Bibliotheken sie ebenfalls enthalten ist. Man erhält eine Liste von Bibliotheken, die man nach der Häufigkeit der Übereinstimmungen absteigend ordnet. Nun berechnet man den Median der gruppierten Häufigkeiten und verwirft alle Bibliotheken, die weniger Übereinstimmungen aufweisen:

|L1 ∩ L2| = 3 |L1 ∩ L3| = 5 |L1 ∩ L4| = 3 |L1 ∩ L5| = 4 |L1 ∩ L6| = 1 |L1 ∩ L7| = 1	Median {5, 4, 3, 1} = 3,5	Im weiteren würden nur noch L3 und L5 betrachtet werden.

Durch die Trunkierung der Liste hält man Bibliotheken mit zu kleinen Überschneidungen aus der weiteren Analyse heraus.

Nach **Hypothese 1** hat man nun diejenigen Bibliothek identifiziert, die im Gesamtsystem die meisten Ressourcen mit L1 teilen.

Beinhaltet die Bibliothek L1 Ressourcen zu verschiedenen Interessengebieten, kann man, statt die gesamte Bibliothek zu betrachten, durch Angabe eines oder mehrer Tags, die Menge der zu untersuchenden Ressourcen im Vorfeld einschränken. Auch könnte man die Relation der Anzahl der Ressourcen in einer dritten Bibliothek Lx zu ihrer Schnittmenge mit L1 betrachten und den Bibliotheken damit ein Ranking zuweisen. Je größer der Anteil der Schnittmenge an der Gesamtgröße ist, desto wahrscheinlicher ist es, dass Lx weitere relevante Ressourcen umfasst. Je kleiner der Anteil ist, desto breiter gestreut ist das Interessengebiet des Besitzers von Lx und desto kleiner ist die Wahrscheinlichkeit, weitere relevante Dokumente zu finden. Diese Aussage ist noch zu überprüfen und stellt lediglich eine Überlegung dar.

2. Nun betrachtet man die Ressourcen aus L1 ∩ L3 und identifiziert die Tags, mit denen der Besitzer von L3 diese Ressourcen getaggt haben. So erhält man eine Liste von Tags, die man nach Häufigkeit des Vorkommens absteigend ordnet, den Median der Häufigkeiten berechnet und diejenigen Tags verwirft, die weniger oft vorkommen. Damit hat man das Vokabular V gefunden, mit dem der Besitzer von L3 die Ressourcen innerhalb der Schnittmenge L1 ∩ L3 getaggt hat.

Nach **Hypothese 2** werden Ressourcen aus L3 \ L1, die denen aus L1 ∩ L3 ähnlich sind, mit Tags versehen worden sein, die in V enthalten sind.

Durch das Erstellen eines Vokabulars für eine Teilmenge einer Bibliothek eines Benutzers umgeht man das Problem unterschiedlicher Tags, die von verschiedenen Benutzern für den gleichen Begriff gewählt wurden. Wie am Beispiel aus dem Video von Udell gezeigt wurde, können zwei Benutzer die gleiche Ressource aus unterschiedlichen Motivationen heraus mit verschiedenen Tags getaggt haben. Indem man für die Suche in einer Bibliothek nur die Folksonomy (das in extrahierte Vokabular ist Teil der Folksonomy des Benutzers) der Bibliothek selbst nutzt, umgeht man das Pro-

blem der Disambiguierung, wenn man davon ausgeht, dass ein Benutzer für den gleichen Begriff immer das gleiche Tag benutzt.

Um sicherzustellen, dass das gefundene Vokabular sinnvoll auf L3 \ L1 anzuwenden ist, sollte darauf geachtet werden, dass es nicht zu wenige Tags beinhaltet. Genauso sollte es nicht zu viele Tags beinhalten. Man könnte also zum Beispiel, statt das Vokabular am Median abzuschneiden, mit relativen Werten arbeiten und V als die 20% des Vokabulars definieren, die am häufigsten vorkommen, wobei V mindestens fünf Tags beinhalten muss. Desweiteren könnte man auch dem Benutzer die Möglichkeit geben, sich selbst aus dem gesamten Vokabular das Vokabular V zusammenzustellen. Damit könnte der Prozess zwar nicht mehr automatisch ablaufen, aber eventuell gewinnt man dadurch bessere Suchergebnisse. Eine weitere Möglichkeit wäre die Benutzung einer Stoppwortliste, in der bekannte, nicht-deskriptive Tags wie „toread“ enthalten sind. Eine solche Liste existiert nicht und müsste im Vorfeld erstellt werden.

3. Im nächsten Schritt betrachtet man die Ressourcen der Menge L3 \ L1 und die Tags, mit denen sie vom Besitzer der Bibliothek L3 getaggt wurden. Man prüft für jede Ressource, wie groß die Übereinstimmung zwischen der benutzen Tag-Gruppe T und dem im vorherigen Schritt erstellten Vokabular V ist. Besteht keine Übereinstimmung, ist diese Ressource nicht interessant. Besteht eine Übereinstimmung, so wird die Mächtigkeit der Übereinstimmung $|V \cap T|$ berechnet.

Man könnte hier eine Grenze einfügen und bestimmen, dass die Mächtigkeit der Übereinstimmung mindestens x betragen muss. x kann zum Beispiel so definiert sein, dass es 50% oder 80% der Anzahl der Elemente aus V beträgt. So kann man strenger oder weniger streng filtern, sollte die Anzahl der Ressourcen der Treffermenge zu groß sein oder zu viele irrelevante Treffer geliefert werden. Die Ressourcen, bei denen $|V \cap T| \geq x$ gilt, erhalten als Score $|V \cap T|$ zugeordnet.

Nach **Hypothese 3** erhält man so eine gewichtete Liste von unbekannten, aber potentiell interessanten Ressourcen.

4. Nun wiederholt man die Schritt 2 und 3 für alle weiteren in Schritt 1 identifizierten Bibliotheken. Trifft man dabei auf eine Ressource, die auch in L3 \ L1 enthalten ist, erhöht man ihren Score entsprechend Schritt 3.

   Nach **Hypothese 4** erhöht man den Score einer Ressource, die in mehreren Bibliotheken auftaucht, um so abzubilden, dass sie potentiell interessanter ist als andere Ressourcen, die nur in einer oder weniger Bibliotheken auftaucht.

Im Ergebnis erhält man eine gewichtete Liste von Ressourcen aus L3 \ L1 und L5 \ L1, die mit Tags aus dem jeweiligen Vokabular getaggt sind und damit ähnliche Inhalte haben, wie die Ressourcen aus L1.

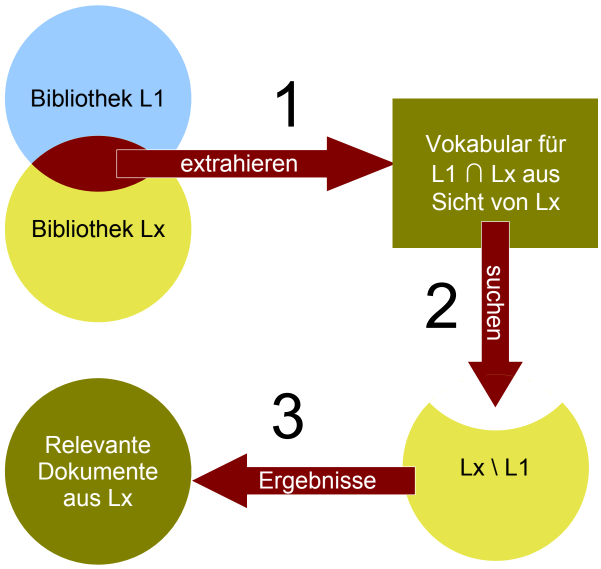


Abbildung 2-13: Darstellung des Ablaufs zur Identifikation interessanter und unbekannter Ressourcen.

Dieses Vorgehen lässt sich für alle Taggingsysteme implementieren, die ein „bag-model" für Tags aufweisen, und die nach dem „free-for-all"-Prinzip aufgebaut sind, zum Beispiel del.icio.us oder CiteULike.

3 Folksonomy

a Definitionen

Der Begriff ist ein Kunstwort, bestehend aus dem englischen Substantiv *folk* und dem griechischen *nomia* im Gegensatz zu Taxonomy, dass aus *taxis*, der Ordnung und *nomia* besteht. Er beschreibt damit eine von Menschen erstellte Ordnung oder Handhabung. Folksonomy möchte im Gegensatz zu Taxonomy deutlich machen, dass die Ordnung nicht von vornherein festgelegt ist, sondern durch Benutzereingaben entsteht.

Folksonomies entstehen im Zuge von Tagging, durch das Tagging. Eine Folksonomy stellt die Gesamtheit aller Tags dar, die Benutzer eines auf Tags basierten Systems bis zu einem bestimmten Zeitpunkt eingegeben haben. So wächst eine Folksonomy durch Nutzung des Taggingsystems und verändert sich sich mit jeder Benutzer-Eingabe.

Im Gegensatz zu professionell gepflegten Systemen gehen Folksonomies formale Aspekte völlig ab. Es fehlen Verweise und Verknüpfungen, Siehe-Auch-Hinweise und ähnliches. Es fehlen auch hierarchische Gliederungen, die Abhängigkeiten zu Ober- und Unterbegriffen deutlich machen.

Ein weiterer Gegensatz ist die grundsätzliche Offenheit von Folksonomies. Glaubt ein Benutzer, dass es sinnvoll wäre, einer Ressource ein weiteres Tag zu geben, benutzt er es einfach und fügt damit der Folksonomy einen Begriff hinzu.

Guy und Tonkin beschreiben Folksonomies daher so:

> *The result [des Taggens] is an uncontrolled and chaotic set of tagging terms that do not support searching as effectively as more controlled vocabularies do.*

Sie führen weiter aus, dass eine Verbesserung des Taggingprozesses zu nutzbaren Tags führen könnte und dadurch den Ansatz eines offenen Vokabulars deutlich aufwerten würde. In Kapitel Tagging Support, S. 34, werden ihre Vorschläge aufgegriffen.

b Struktur entsteht durch Wachstum

Bei Folksonomies, den Taxonomien, die durch Tagging enstehen, lässt sich nach einer gewissen Zeit der Nutzung eine Struktur der Tags ablesen. So können recht schnell die am häufigsten genutzten Tags hervorgehoben werden oder auch solche, die häufig mit einander verwendet werden.

Diese Listen sind nicht statisch zu sehen, sondern dynamisch in der Zeit. Sie ändern sich im Laufe von Tagen, manchmal nur Stunden. Solche Abbildungen können daher immer nur Schnappschüsse sein. Folksonomies bieten weder Hierarchien noch Werkzeuge bekannter Ontologien oder Thesauri. Tatsächlich bilden sie immer nur in einen Ist-Zustand aller im System benutzten Tags ab, ohne diese in Beziehungen zu setzen, wie kontrollierte Vokabulare dies tun. Natürlich sind Aussagen über die am häufigsten genutzten Tags oder die am häufigsten gemeinsam vorkommenden Tags möglich, von reinen Anzahlen („Wie oft wurde Tag x benutzt?“) einmal abgesehen.

In den meisten Systemen ist es nicht möglich, sprachlich zu unterscheiden, weshalb alle Tags in der selben Folksonomie landen, ohne Unterscheidung nach Sprache oder Einsatzzweck. Die meisten Systeme führen selbst auch kein Stemming durch oder bereiten ihre Tags anders auf eine halb intelligente Weiterverarbeitung vor. Tatsächlich wird jedes Wort der natürlichen Sprache, das technisch innerhalb des Regelsystems als Tag darstellbar ist, als solches in die Folksonomie des Systems aufgenommen.

c Folksonomies und das Power Law

Golder und Huberman, und andere bemerken, dass die Häufigkeitsverteilung von Tags in einer Folksonomy einem Power Law[33] folgt.

33 Siehe http://de.wikipedia.org/wiki/Power_law – Zugegriffen am 07.08.2006.

Power Laws beschreiben eine in vielen Gebieten nachgewiesene Verteilung von Mengen. Dabei geht man von exponentiellem Wachstum sowohl der Menge der beobachteten Objekte als auch ihrer Größe aus. So sind zum Beispiel die Städte und Ortschaften eines Landes, absteigend nach ihrer Größe geordnet, nach den Power Laws verteilt. Die Pareto-Verteilung und Zipfs Gesetz sind weitere Beispiele für Power Laws.

Im Zusammenhang mit Tagging und Folksonomies wird häufig anstatt von einer Verteilung nach den Power Laws vom sogenannten „Long tail“ (vgl. [Anderson2004]) gesprochen. Damit ist die Vielzahl der unterschiedlichen Tags und ihre Häufigkeit gemeint, mit der die Gesamtheit der Benutzer eine Ressource taggen. Hier ein Beispiel dafür, an dem im folgenden die Bedeutung des „Long tail“ erläutert wird:

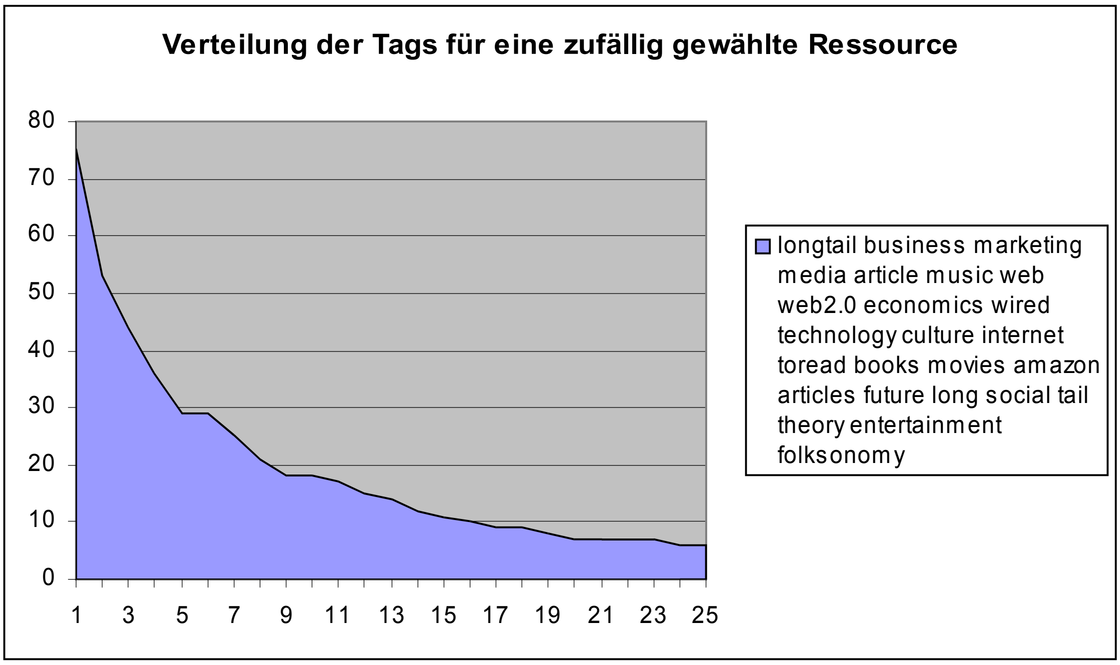


Abbildung 3-1: Verteilung der Tags für eine zufällig ausgewählte Ressource.

Die Prinzipien hinter dem „Long tail“-Konzept, das seine Wurzeln in der Ökonomie hat, lauten nach Anderson:

1. Make everything available.

2. Cut prize in half. Now lower it.

3. Help me find it.

Anderson, der dieses Konzept zum ersten Mal formulierte, argumentiert, dass man zwar mit 20% seiner Waren 80% des Gewinns macht, aber man darüber die übrigen 80% der Waren nicht vernachlässigen sollte. Übertragen auf Folksonomies bedeutet das, dass 20% der Tags für 80% aller Ressourcen benutzt werden. Diese Prozentangaben sind natürlich nur Näherungswerte und nicht wörtlich zu verstehen. Die Verbindung zu Folksonomies ist die Offenheit der Folksonomies. Sie erlauben jedes mit einer Tastatur eingebbare „Wort“ als Tag, ohne dabei den Nutzer in irgendeiner Art und Weise einzuschränken. Wie es dann dazu kommt, dass einzelne Tags so exponentiell häufiger vorkommen, liegt in der Natur von del.icio.us und allen Taggingsystemen begründet: sie sind nach den Prinzipien des „bag-models“ (siehe Abschnitt Aggregation, S. 35) aufgebaut und erlauben jedem Nutzer, jede Ressource zu taggen („free-for-all“, siehe Abschnitt Tagging Rights, S. 31).

Man betrachtet im Fall von Folksonomies also die Häufigkeiten, mit der die einzelnen Tags vorkommen und vergleicht sie. Daraus folgen Diagramme wie in den folgenden Abbildungen. Sie wurden aus Daten erzeugt, die am 25.07.2006 von 14.00h bis 14.30h für zwei zufällig ausgewählte Benutzer von del.icio.us per RSS-Feed gewonnen wurden. Die Daten beinhalten alle Ressourcen, die jeder der vier Benutzer in seiner Link-Sammlung hat, sowie die dazu genutzten Tags. Durch Gruppierung der Tags und der Auswertung, welches Tag wie häufig im jeweiligen Datensatz vorkommt, konnten diese Diagramme erstellt werden.

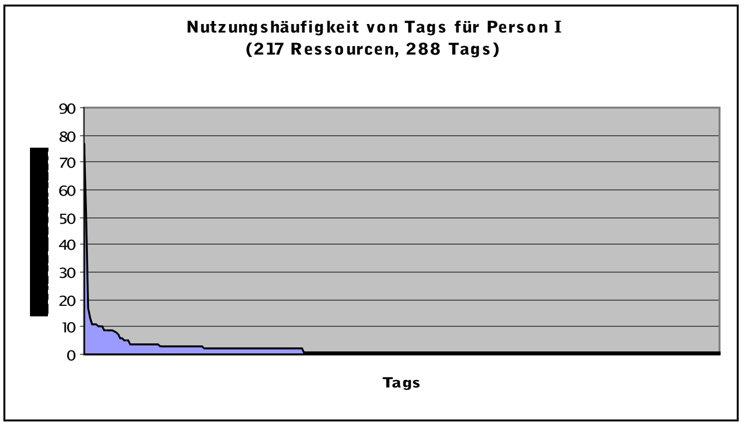


Abbildung 3-2: Häufigkeitsverteilung der Tags eines Nutzers von del.icio.us.

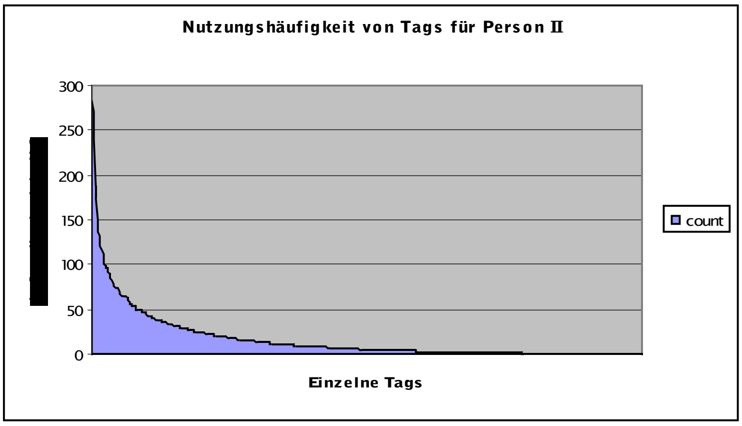


Abbildung 3-3: Häufigkeitsverteilung der Tags eines Nutzers von del.icio.us.

d Visualisierung der Struktur

i Tag Clouds

Meist werden Folksonomien per Tag Clouds visualisiert. Dabei nimmt man ein Subset von Tags aus der Menge aller, meist die 100 häufigsten, ordnet sie nach dem Alphabet und vergrößert dann einzelne Tags je nach Nutzungshäufigkeit, so dass man beispielsweise die häufigsten zwanzig deutlich vergrößert darstellt. Statt durch Vergrößerung ist der Effekt auch mit Farben zu machen, wobei man auf eine Vergrößerung meist nicht verzichtet. Del.i-cio.us nutzt die Farbdimension, um solche Tags zu markieren, die man selbst bereits benutzt hat:

Die in rot gedruckten Tags sind solche, die man selbst bereits benutzt hat. Wie zu sehen ist, liegt der Fokus der bei del.icio.us vornehmlich verwalteten Bookmarks auf technischen, internet-relevanten Themen.

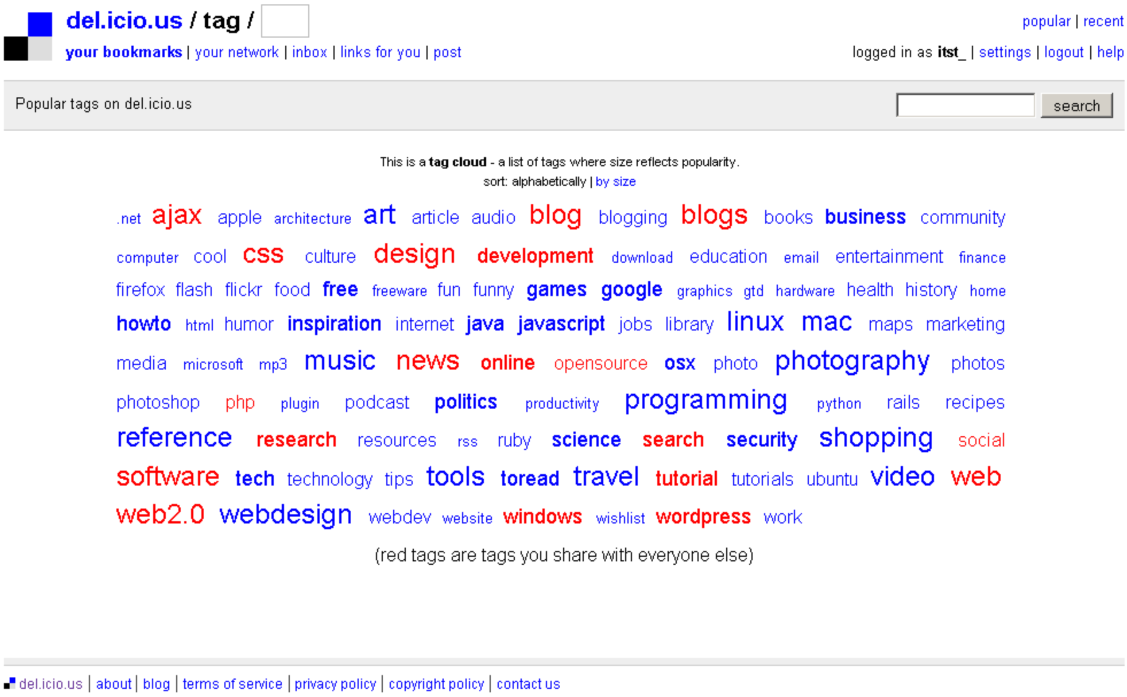


Abbildung 3-4: Screenshot der Tag Cloud von del.icio.us vom 11.05.2006.

Im Moment sieht man immer wieder Beispiele für solche Tag Clouds an Stellen, bei denen man erkennt, dass weder Tags noch Folksonomies die Basis für die Abbildungen sein können. So werden Tag Cloud-ähnliche Designs als Navigationselemente für Websites benutzt, zum Beispiel bei McDonalds oder Yahoo! Tech:



Abbildung 3-6: Navigation auf http://ww.mcdonalds.de/ - Dargestellt als Tag Cloud.

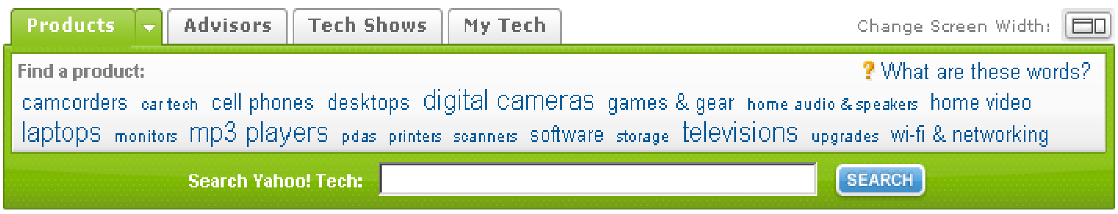


Abbildung 3-7: Navigation auf http://tech.yahoo.com/ - Dargestellt als Tag Cloud

ii Interaktive Visualisierungen von Folksonomies

Solche und andere Darstellungen von Folksonomies bedienen sich meist einer interaktiven Grafik mittels Flash- oder Java-Programmen. Sie erlauben, die eigene Folksonomy visuell darzustellen und Zusammenhänge sichtbar zu machen.

Delicious Soup

Dieses[34] Werkzeug setzt einzelne Tags in Beziehung und zeigt so, welche Tags häufig mit anderen gemeinsam eine Tag-Gruppe bilden. Es visualisiert die Folksonomy eines einzelnen Benutzers auf Basis seiner Linksammlung bei del.icio.us.

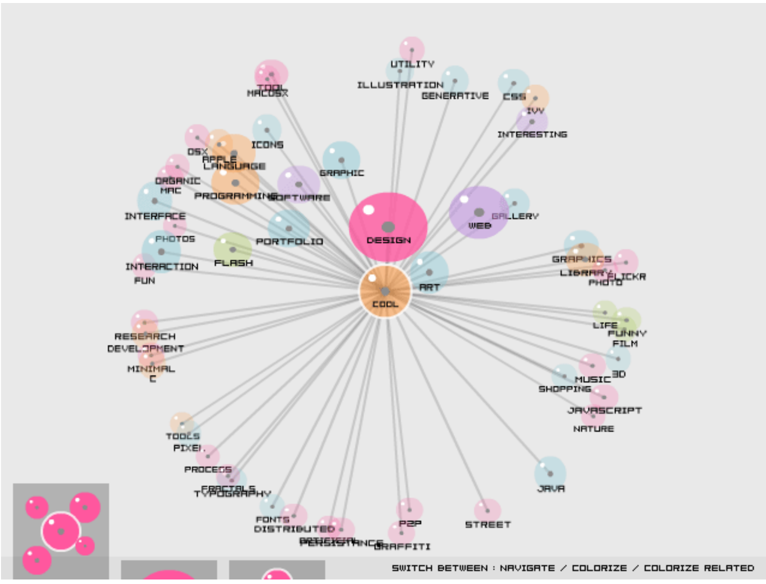


Abbildung 3-8: Darstellung der persönlichen Tags mit Delicious Soup

Delicious Soup zeigt anhand der Nähe zweier Tags, dass sie oft gemeinsam vorkommen und anhand der Größe, wie oft ein Tag insgesamt vorkommt. An dem folgenden Beispiel wird dies deutlich, in dessen Zentrum das Tag „api" steht. Der Mauszeiger steht über dem Tag „tags" (im Screenshot ist der Mauszeiger nicht zu sehen).

Der Text am unteren Bildrand von Abbildung 3-9 lautet sinngemäß: tags (used 54 times) :: since 225 posts :: has 23 related (has 2 relations with api). Hierzu einige kurze Erläuterungen:

34 Siehe http://www.zitvogel.com/delicioussoup/ - Zugegriffen am 04.06.2006.

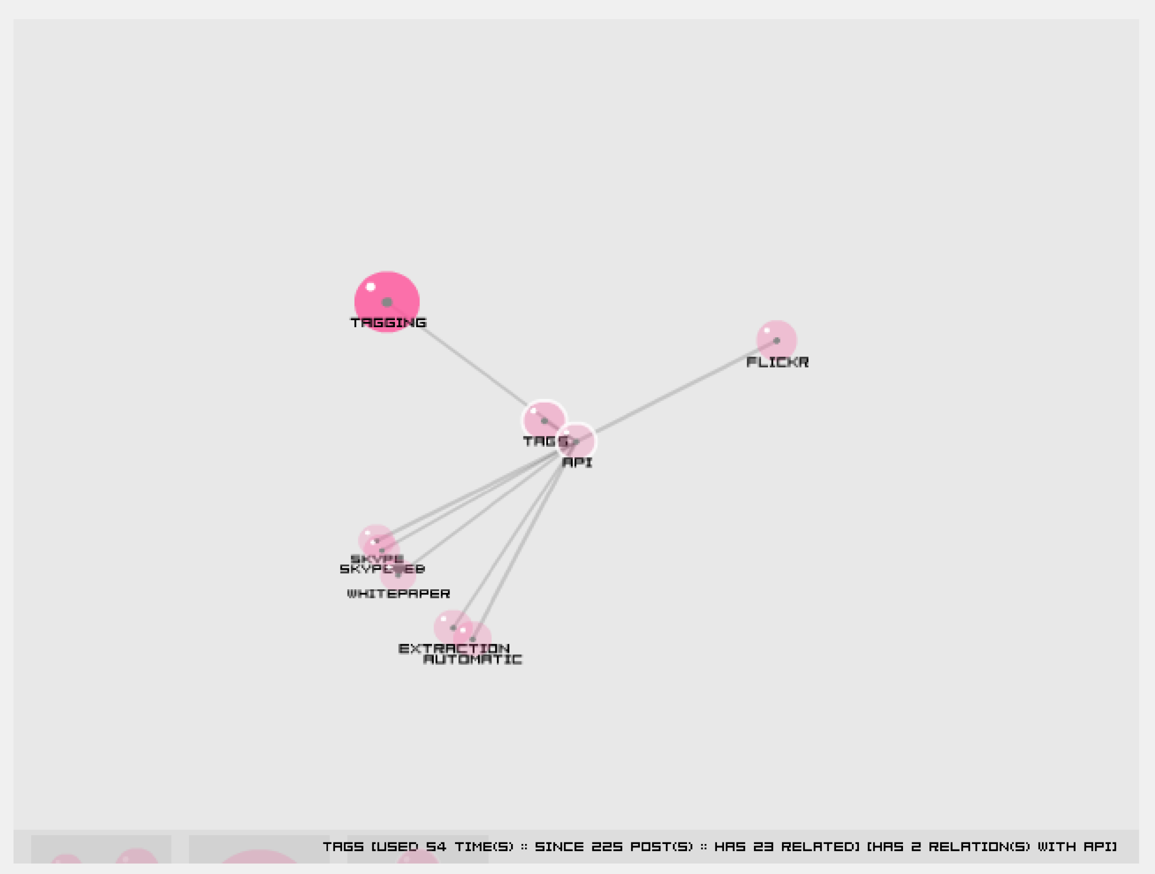


Abbildung 3-9: Screenshot von Delicious Soup. Er zeigt die Visualisierung der Beziehungen des Tags "api".

1. „tags (used 54 times)" - gibt Aufschluss über die Frage, wie oft dieses Tag in der untersuchten Folksonomy vorkommt.

2. „since 225 posts" - diese Angabe entspricht dem Alter des Tags in der Folksonomy. Kennt man die Anzahl der Ressourcen der zugrunde liegenden Linksammlung, kann man mit der Angabe der Nutzungshäufigkeit eine näherungsweise Aussage über die Nutzungsfrequenz machen: teilt man das Alter eines Tags (gemessen in „posts", also getaggten Ressourcen) durch die Nutzungshäufigkeit, erhält man die Frequenz, mit der dieses Tag benutzt wird. Hier also: 225 : 54 = 4,17. Das Tag „tags" wird also durchschnittlich für jede vierte Ressource der 225 Ressourcen genutzt, die seit dem ersten Auftreten des Tag getaggt wurden.

3. „has 23 related" - das Tag hat seinerseits Verbindungen zu 23 anderen Tags ,taucht also mit ihnen gemeinsam in verschiedenen Tag-Gruppen auf.

4. „has 2 relations with 'api'" - es gibt zwei Ressourcen in der Linksammlung, die mit „api" und „tags" getaggt wurden.

Delicious Soup ist als Hilfsmittel zu sehen, das dem Benutzer Einblicke in die Folksonomy seiner Linksammlung gibt. Durch die interaktive Darstellungsweise kann er durch Mausklicks seine Tags „entdecken" und Zusammenhänge leichter erschließen, als dies durch eine Tag Cloud möglich wäre.

Ravelicous

Ravelicous[35] stellt drei Visualisierungen zur Verfügung, SpaceNav, TagClouds und Grouper. Es benutzt als Basis die Folksonomy eines einzelnen Benutzern von del.icio.us.

SpaceNav ist eine spielerische Visualisierung, die zur Darstellung der Beziehungen einzelner Tags dient. Es repräsentiert einzelne Tags als Kreise. Je größer der Kreis, desto häufiger wurde das Tag innerhalb der Folksonomy benutzt.

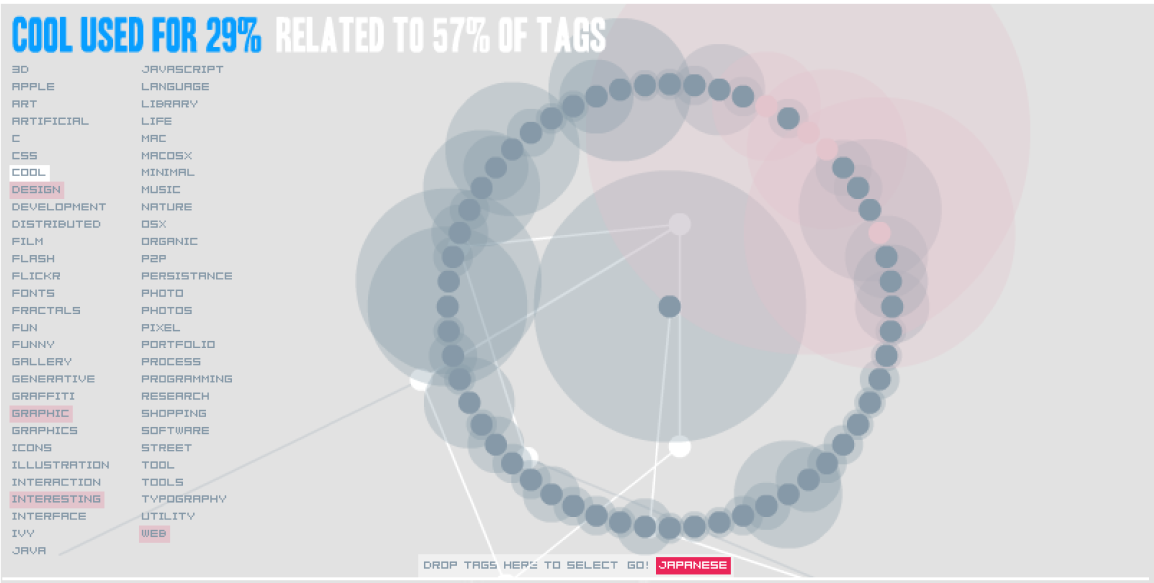


Abbildung 3-10: Visualiserung der Beziehungen des Tags "cool" mit Ravelicous SpaceNav.

35 Siehe http://www.ivy.fr/revealicious/ - Zugegriffen am 04.05.2006.

Im Hintergrund des Screenshots aus Abbildung 3-10 sieht man weiße Kreise und sie verbindende Linien. Dies ist eine Besonderheit von SpaceNav: die Darstellung des Klick-Pfades („path“), den man durch die Visualisierung bisher genommen hat. Klickt man auf die Repräsentation eines Tags in der rechten Bildhälfte, wechselt man zur Darstellung der Beziehungen dieses Tags. SpaceNav erstellt dann eine weiße Linie, die dieses neugewählte Tag mit dem vorherigen verbindet.

TagClouds stellt die Tags einer Folksonomy in einer ähnlichen Form dar, wie statische Tag Clouds das tun. Durch die Interaktivität jedoch gewinnt die Darstellung an Informationswert. In der Darstellung werden die am häufigsten benutzten Tags mit einem weißen Hintergrund versehen. Tags ohne Hintergrundfarbe werden im Vergleich weniger oft und hellere Tags kaum benutzt.



Abbildung 3-11: Visualierung der Beziehungen eines Tags mit TagClouds.

Ausgewählt ist hier das Tag „usability“. Es steht mit den Tags mit rotem Hintergrund in Beziehung. Je heller die rote Färbung, desto weniger oft kommt „usability“ mit diesem Tag in einer Tag-Gruppe vor. Oben rechts ist das Tag „cool“ mit einem weißen Hintergrund versehen. Dies zeigt an, dass dieses Tag sehr oft in der Folksonomy vorkommt.

Die Visualisierung mit TagClouds bietet darüber hinaus die Möglichkeit, mehrere Tags in die Textzeile am unteren Bildrand zu ziehen und dann alle Ressourcen anzuzeigen, die mit diesen Tags getaggt wurden.



Abbildung 3-12: Die gezoomte Darstellung einer Folksonomy mit TagClouds stellt den Text viel zu klein dar und macht ihn so unleserlich.

Die in den beispielhaften Screenshots gezeigten Folksonomies sind recht klein, bestehen also aus nur wenigen Tags. Ist die Folksonomy größer, wird die Navigation unübersichtlich, weil nicht mehr die gesamte Folksonomy im Navigationsbereich sichtbar ist, sondern nur noch ein Teil. Zwar kann man die Darstellung verkleinern und vergrößern, doch wird die Schrift dann so klein, dass man nichts mehr erkennen kann.

Grouper ist ein Versuch eines interaktiven Folksonomy-Managers. Außerdem stellt es ähnliche Informationen über einzelne Tags und ihre Beziehungen bereit, wie SpaceNav und TagClouds es tun.

Hier wurde das Tag „bookmarking" ausgewählt und mit Grouper automatisch die mit ihm in Beziehung stehenden Tags zu einer Gruppe zusammengefasst.

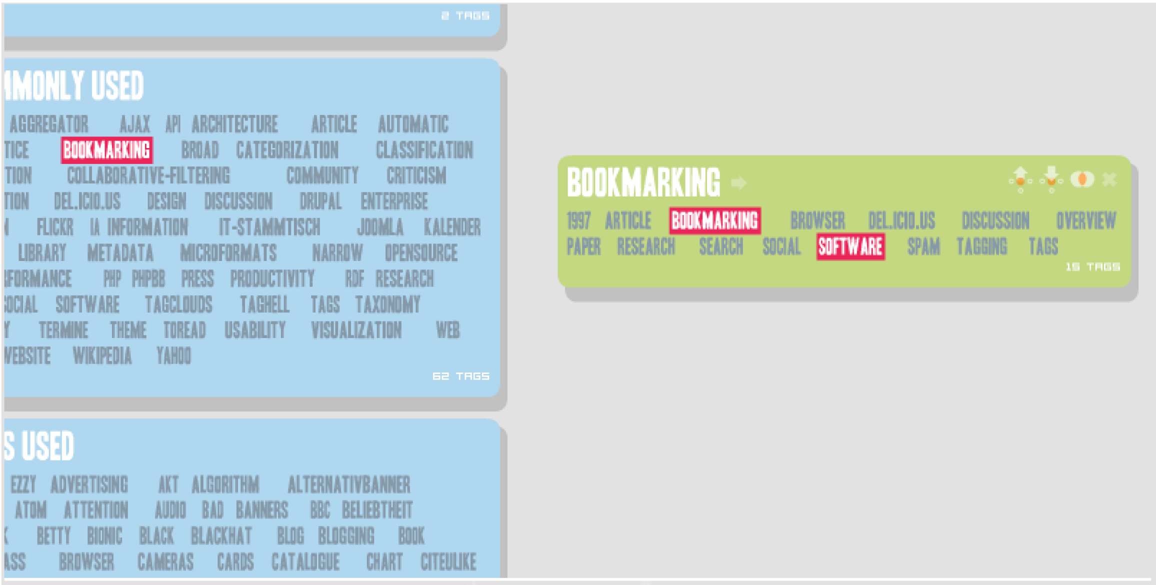


Abbildung 3-13: Grouper bietet die Möglichkeit, Tags nach Gruppen zu ordnen. Wie TagClouds ist es mit größeren Folksonomies überfordert, was den Raum der Anzeige (Viewport) angeht.

iii Touchgraph-Visualisierungen

Touchgraph[36] ist ein in der Programmiersprache Java geschriebenes Framework zur graphischen Darstellung von Netzwerken jeglicher Art. Eaton hat mittels dieses Frameworks ein Visualierungswerkzeug erstellt, das Verhältnisse zwischen Tags abbilden kann. Sein Werkzeug betrachtet dabei die gesamte Folksonomy von del.icio.us.

Abbildung 3-14 zeigt die Verbindungen der Tags „tagging" und „folksonomy". Zunächst wurden die Relationen für „tagging" dargestellt. Eine der Relationen von „tagging" ist „folksonomy". Durch einen Doppelklick auf „folksonomy" spannte sich auch dessen Beziehungsnetzwerk auf.

Eine weitere Abbildung[37] zeigt das Beziehungsnetzwerk von „socialsoftware" mit weiteren aktivierten Netzwerken. Neben „socialsoftware" werden die Netzwerke von „wiki", „web", „music", „del.icio.us", „folksonomy" und weiteren Tags geöffnet. Touchgraph verknüpft dabei bereits angezeigte Tags mit denen aus neu aktivierten Netzwerken, so dass eine Gesamtübersicht der Beziehungen besteht. So gehört „ontology" zum Netzwerk von „meta-

36 Siehe http://touchgraph.sourceforge.net/ - Zugegriffen am 15.05.2006.
37 Siehe http://many.corante.com/archives/2005/01/26/visualizing_the_collective_brain.php – Zugegriffen am 15.05.2006.

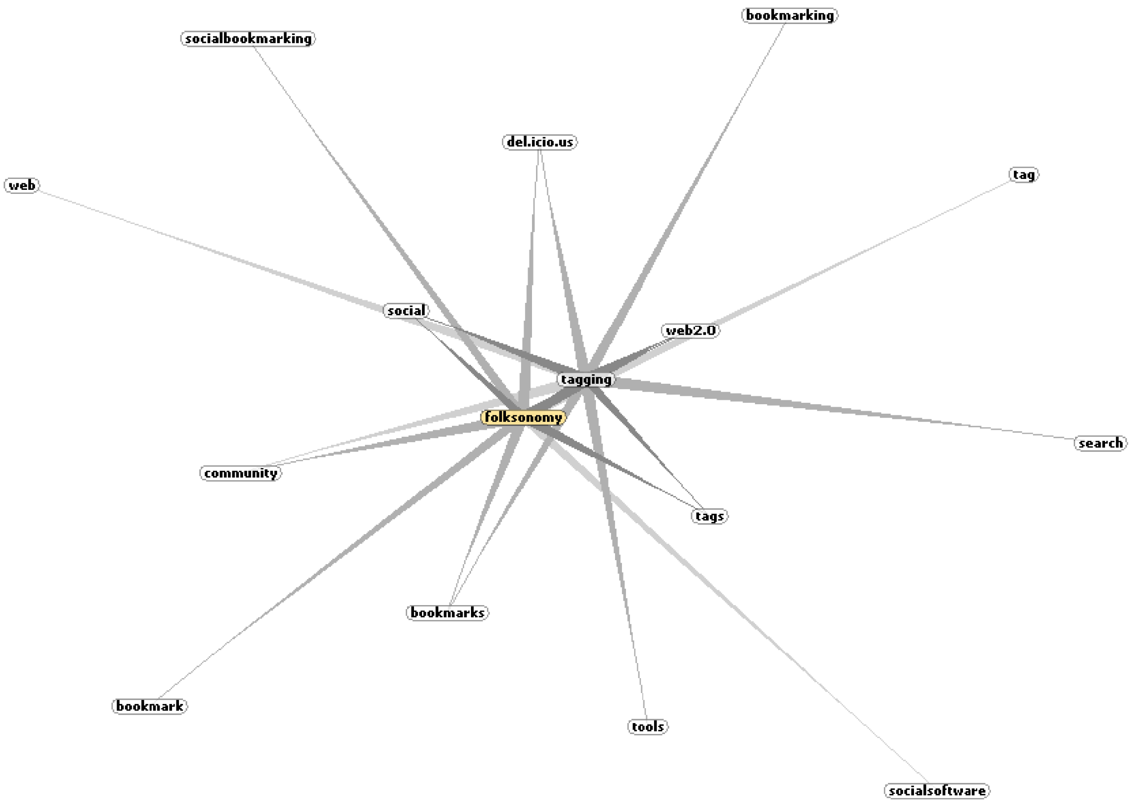


Abbildung 3-14: Visualisierung der Beziehungen der Tags "tagging" und "folksonomy" mit Touchgraph.

data“ welches wiederum zum Netzwerk von „folksonomy“ gehört. Zu seinem eigenen Netzwerk gehört „web“, was auch Teil des Netzwerks von „socialsoftware“ ist.

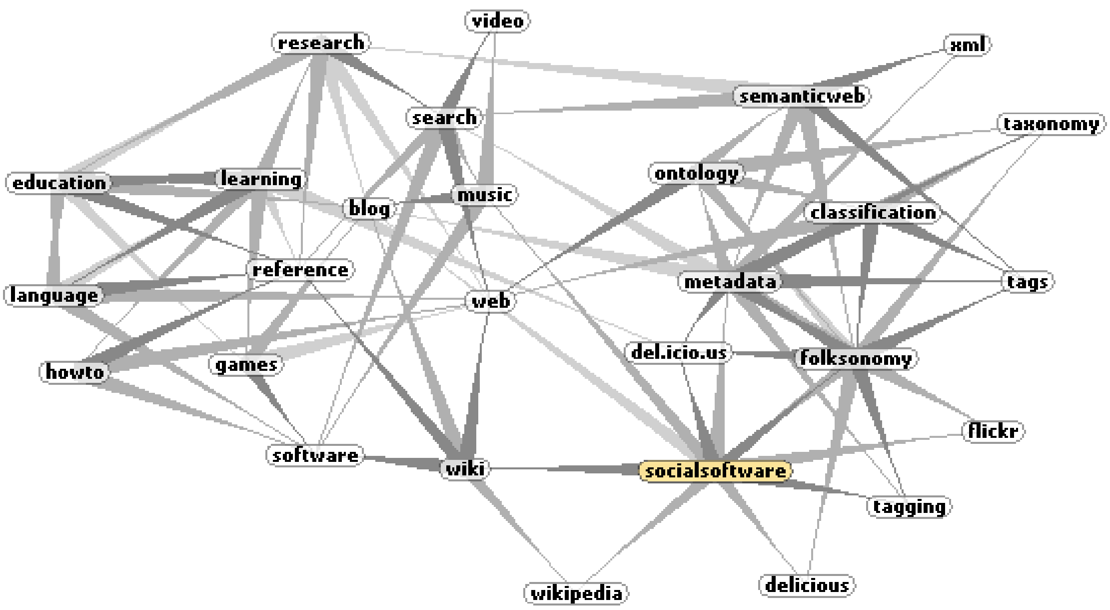


Abbildung 3-15: Visualisierung der Beziehungen rund um "socialsoftware".

4 Kritik

Noch fehlen Folksonomies die Möglichkeiten von Thesauren und Ontologien. Mit den Arbeiten an Tag-Clustern (siehe Abschnitt Tag-Cluster, S. 80f) könnte sich diese Kritik bald als überholt betrachten lassen. Den meisten Klassifikationssystemen fehlt aber die Einfachheit der Nutzung, die Folksonomies bereitstellen. Der Benutzung zur Indexierung der Colon-Klassifikation beispielsweise geht ein Studium der Regeln voraus, so dass kaum ein Endnutzer sie nutzen würde, um zum Beispiel seine Bookmarks zu verwalten. Auch ist das Ergebnis der Indexierung nach Colon für den Laien kaum verständlich. Die Ressource „Diagnostik der Lungentuberkulose in Frankreich 1989“ würde im Ergebnis zu indexiert sein: „L, 45:421:3.53.N89“. Dies gilt auch für die Universal Decial Classification und weitere Klassifikationssysteme, die mit Trennzeichen unterschiedliche Beziehungen zwischen den Termen herstellen. Klassifikationssysteme und kontrollierte Vobabularien wachsen nur langsam und das Hinzufügen neuer Begrifflichkeiten bedarf Genauigkeit und sehr gute Kenntnisse des Systems. An der Dewey-Dezimalklassifikation wird dies besonders deutlich. Die Festlegung auf Breite und Tiefe einer solchen Klassifikation behindert ihr zukünftiges Wachstum und macht Änderungen fast gänzlich unmöglich. Doch natürlich haben diese Systeme auch Vorteile gegenüber Folksonomies.

Das Problem der Disambiguierung stellt sich regelmäßig in Folksonomies. Neben dem erwähnten Beispiel im Abschnitt Homonyme und Synonyme, S. 19, an dieser Stelle einige weitere, die diesen Aspekt näher erläutern. Der Begriff „queso“ steht in der spanischen Sprache für „Käse“. Betrachtet man die Ressourcen, die mit diesem Begriff bei del.icio.us getaggt wurden[38], fällt auf, das sich die Bedeutung dieses Begriff innerhalb der Folksonomy von del.icio.us im Laufe der Zeit wandelte. Torres (vgl. [Torraumes2006])hat dies getan und kommt zu dem Schluss, dass vor dem 17.06.2006 mit dem Tag „queso“ der spanische Begriff „Käse“ gemeint war. Nach dem 17.06.2006 allerdings taucht das Tag immer wieder in Zusammenhang mit „Queso - a Semantic Web/Web 2.0 server“[39] auf, einem Atom[40]-Server. In

38 Siehe http://del.icio.us/tag/queso – Zugegriffen am 09.08.2006.
39 Siehe http://torrez.us/archives/2006/07/17/471/ - Zugegriffen am 09.08.2006.
40 Siehe http://www.atomenabled.org/ – Zugegriffen am 09.08.2006. Atom stellt ein Format für Content Syndication und ein Protokoll zum Datenaustausch im Rahmen des Atom-Formats bereit.

einem kontrollierten Vokabular könnte dies nicht geschehen, da diese Systeme extrem präzise und unmissverständlich sind. Vor allem die Synonym- und Homonymkontrolle ist eine Stärke kontrollierter Vokabularien/Thesauri gegenüber Folksonomies.

Wie am Beispiel der Sonderzeichen in Tags, S. 22, deutlich wurde, können Folksonomies eine gewisse Ähnlichkeit mit Facettenklassifikationen haben. Dabei ist jedoch zu beachten, dass auch die Befürworter dieser Praxis immer dazu raten, auch und vor allem allgemeine, das heißt nicht mit Sonderzeichen versehene, Tags zu benutzen. Die Gefahr, sich hier mit Sonderzeichen und dergleichen zu verzetteln, ist natürlich dennoch gegeben. Es ist jedoch davon auszugehen, dass die Masse der Benutzer nicht zu dieser Praxis übergeht, schlicht weil sie – als informationswissenschaftliche Laien – kein Wissen über solche Methoden haben.

a Vorteile von Folksonomies

Im Gegensatz zu kontrollierten Vokabularen und Klassifikationssystemen ist der Einstieg in eine Folksonomy für den Endnutzer wesentlich einfacher. Auch das Konzept des Taggens, der Grundlage von Folksonomies, ist leichter zu verstehen und aufzunehmen – abgesehen von der Tatsache, dass es nicht der Endnutzer ist, der Terme aus einem kontrollierten Vokabular auf eine Dokumentarische Beschreibungseinheit anwendet, sondern ein Informationsspezialist mit Fachausbildung. Die Einstiegshürde ist geringer, die Lernkurve weniger steil. Während Klassifikationssysteme vor allem als Hilfsmittel zur Suche eingesetzt werden („Ich suche Literatur über den Zusammenbruch der UdSSR im Zusammenhang mit der Wiedervereinigung Deutschlands sowie der Gewerkschaftsbewegung in Polen.“), dienen Folksonomies vor allem dem Browsing, dass heißt dem Entdecken von Information. Folksonomies bilden auch die Sprunghaftigkeit und Schnelligkeit besser ab, die dem World Wide Web anhaftet. Durch die Gemeinschaftlichkeit bei der „Erstellung“ von Folksonomies entwickeln sie sich unabhängig von Einzelnen und Einzelinteressen. Durch die Art und Weise der „Erstellung“ von Folksonomies erhalten die Nutzer ein direktes Feedback auf ihre Indexierung einer Ressource und können so weitere Ressourcen und Personen entdecken, die sie interessieren könnten. Folksonomies sind vor allem sind als Werkzeug zur Filterung und Entdeckung von Ressourcen zu sehen.

b Nachteile von Folksonomies

Wie im Abschnitt Probleme, S. 19ff, beschrieben, ergeben sich aus der Freiheit, die der Nutzer im Taggingprozess genießt, einige Nachteile. Zunächst einmal das Problem der Ambiguität von Tags. Unterschiedliche Schreibweisen und unterschiedliche Bedeutungen gleicher Tags machen es schwer, einzelne Tags in einen Kontext einzuordnen. Andererseits kann man, durch Betrachtung der Tag-Gruppe, zu der ein Tag gehört, sehr wohl Rückschlüsse auf die tatsächliche Bedeutung des Tags schließen. Gleichzeitig zeigen erste Ansätze (siehe Abschnitt Aus Folksonomies hierarchische Strukturen bilden, S. 78), dass es durchaus möglich ist, aus Folksonomies Cluster und damit ähnliche Strukturen wie in einem Thesaurus zu bilden. Mit diesen Methoden könnte es möglich sein, aus den flachen Folksonomies polyhierarchische Strukturen zu bilden. Allerdings müsste man diesen Vorgang regelmäßig wiederholen, um keine wichtigen Veränderungen in der Folksonomy zu verpassen. Problematisch ist auch die Schreibung zusammengesetzter Begriffe zu sehen, bei der sich noch kein Defacto-Standard entwickelt hat. Jedoch ist davon auszugehen, dass dies in der nächsten Zeit geschieht. Interessant ist der Ansatz von Technorati, das jegliche Trennzeichen entfernt. So spielt es keine Rolle, ob man nach „opensource“, „open source“, „open-source“ oder „open_source“ sucht[41].

41 Siehe http://technorati.com/tag/ - Zugegriffen am 09.08.2006.Die leicht unterschiedlichen Ergebniszahlen der Suchen nach den verschiedenen Schreibweisen für „open source“ ergeben sich, da Technorati diese Funktion erst nach Einführung der Tag-Unterstützung implementierte.

5 Taggingsysteme am Beispiel von del.icio.us

Das bereits mehrfach erwähnte del.icio.us nennt sich selbst „del.icio.us – social bookmarking“. Es stellt das größte auf Tagging basierte System dieser Art dar und ist seit Ende 2003 öffentlich zugänglich. Im Folgenden wird del.icio.us in die Systematik von Taggingsystemen eingeordnet und der Ablauf des Taggingprozesses beschreiben. Schließlich wird die Navigation in del.icio.us erläutert und der Nutzen dieser Art von System dargestellt.

a „global suggestive free-for-all bag-model“

Ausgehend von der im Abschnitt Ausprägungen von Taggingsystemen, S. 30ff, erläuterten Systematik ordnet sich del.icio.us wie folgt in diese Ordnung ein.

Dimension	Main categories	del.icio.us
Tagging Rights	self-tagging, permission-based, free-for-all	**free-for-all**
Tagging Support	blind, suggestive, viewable	**suggestive**
Aggregation model	bag, set	**bag**
Object type	textual, non-textual	**textual & non-textual**
Source of material	user-contributed, system, global	**global**
Resource connectivity	links, groups, none	**none**
Social connectivity	links, groups, none	**links & groups**

Tabelle 3: Einordnung von del.icio.us in die Dimensionen des Designs von Taggingsystemen nach Marlow et al.

Tagging Rights – del.icio.us erlaubt es jedem Nutzer, jede Ressource im System sowie beliebige neue Ressourcen zu taggen und dem System hinzuzufügen.

Tagging Support – Bei der Indexierung eines neuen Dokuments unterstützt del.icio.us den Nutzer durch Vorschläge für Tags. Dabei zeigt es ihm nicht nur Vorschläge aus den Tags des Nutzers selbst, sondern auch solche Tags, die häufig von anderen Nutzern für diese

Ressource vergeben wurden. Ist die Ressource dem System bislang nicht bekannt, fallen diese Vorschläge natürlich weg.

Aggregation model – del.icio.us erlaubt jedem Nutzer, für jede Ressource eine eigene Tag-Gruppe anzulegen und zu pflegen. Es erlaubt also mehrere Tag-Gruppen pro Ressource, genauer, so viele Tag-Gruppen, wie Nutzer die Ressource in ihre Sammlung aufgenommen haben.

Object type – Prinzipiell kann jedes Dokument, dass per HTTP erreichbar ist, in del.icio.us aufgenommen werden. Es gibt keinerlei Einschränkungen nach Inhalt oder Form des Dokuments.

Source of material – Wie aus dem obigen Absatz deutlich wird, kann in del.icio.us jedes beliebige Dokument getaggt werden, so lange es per HTTP erreichbar ist. Dabei spielt es aus Sicht von del.icio.us keine Rolle, wer der Urheber der Ressource ist.

Resource connectivity – Um Vorschläge für Tags zu machen, stellt del.icio.us lediglich Verbindungen zwischen den Tag-Gruppen her, mit denen eine Ressource bereits getaggt wurde, nicht aber über mehrere Ressourcen hinweg.

Social connectivity – del.icio.us erlaubt es Nutzern, sich gegenseitig zu beobachten und Netzwerke zu bilden.

b Der Taggingprozess

Grundsätzlich gibt es drei Möglichkeiten, seiner del.icio.us-Sammlung eine Ressource hinzufügen: über die del.icio.us-Website direkt, über das von del.icio.us bereitgestellte Bookmarklet[42] sowie über externe Applikationen. Im Folgenden werden vor allem die beiden ersten Möglichkeiten betrachtet, wobei der Prozess an sich der gleiche ist; lediglich der Ausgangspunkt ändert sich.

42 Ein Bookmarklet ist eine Javascript-Anweisung, die man wie einen Bookmark ablegen und bei Bedarf aktivieren kann. Zumeist werden Bookmarklets dazu eingesetzt, die gerade im Browser aktive Seite zu verarbeiten, weiterzuleiten oder umzugestalten.

Benutzt man das Bookmarklet von del.icio.us, so kann man, in dem man es aufruft, die gerade im Browser geladene Seite zu del.icio.us weiterleiten und taggen. Nutzt man direkt die del.icio.us-Seite, muss man die URL der Webseite selbst in das Formular eingeben. Darüberhinaus übernimmt das Bookmarklet auch den Titel der Seite, den man bei dem Weg über das URL-Formular selbst eingeben muss. Für aktive del.icio.us-Nutzer ist das Bookmarklet auf jeden Fall das Mittel der Wahl.

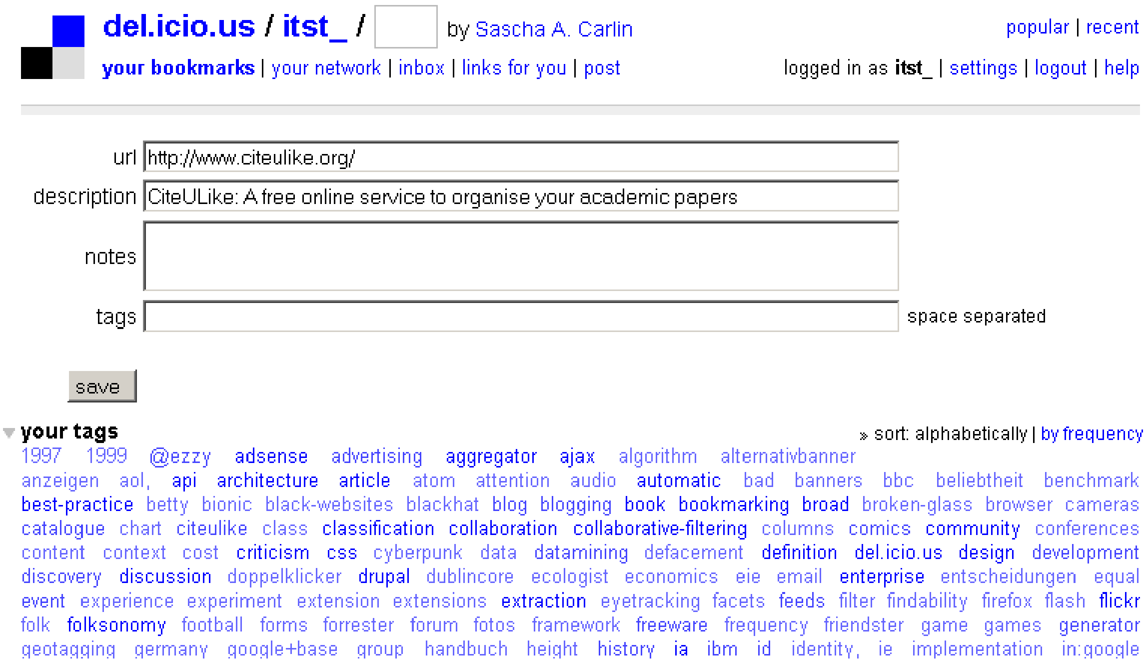


Abbildung 5-1: Ausschnitt aus dem Tagging-Formular von del.icio.us.

Die im Browser geladene Webseite hat die URL http://www.citeulike.org/. Durch einen Klick auf das del.icio.us-Bookmarklet gelangt man zu diesem Formular, in dem URL und Titel der Seite bereits eingetragen sind. Nun hat man die Möglichkeit eine Notiz zu diesem Bookmark anzulegen sowie Tags zu vergeben. Zu sehen sind auch alle eigenen Tags, die man per Mausklick dem neuen Bookmark zuordnen kann. Desweiteren kann man Tags auch direkt in das Eingabefeld eingeben.

Bis zum 24.07.2006 zeigte del.icio.us in diesem Formular auch Tags an, die als „recommended tags" gekennzeichnet waren. Dieses Feature ist im Moment aber nicht aktiviert. Im HTML-Quellcode des Formulars sind die entsprechenden Angaben zwar enthalten, sie wer-

den aber nicht angezeigt. Wie es scheint, liegt ein technisches Problem vor. Leider gibt es von del.icio.us oder Yahoo! Inc. keine Informationen dazu.

Die vorgeschlagenen Tags werden aus den Tag-Gruppen extrahiert, die bereits für die URL angelegt wurden. Der exakte Algorithmus ist leider nicht öffentlich dokumentiert, nach meinen Beobachtungen aber werden die 10-15 am häufigsten benutzten Tags dort angezeigt.

Beginnt man nun damit, Eingaben und das Feld „tags“ zu machen, zeigt del.icio.us direkt diejenigen eigenen Tags an, die mit den gleichen Zeichen beginnen wie die bereits eingegebenen:

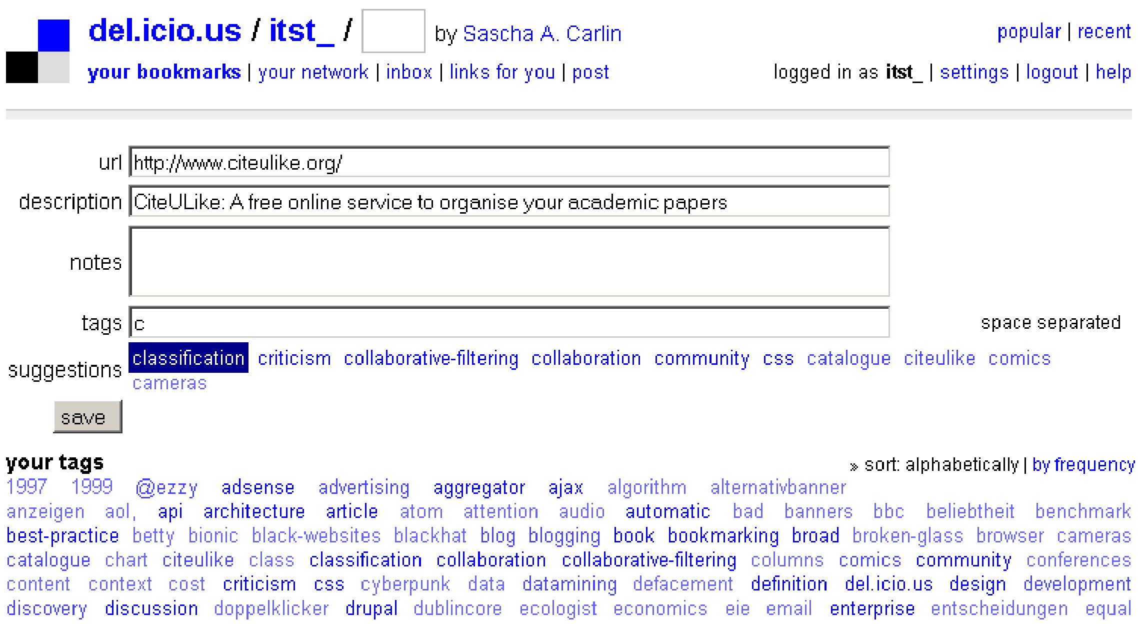


Abbildung 5-2: Type-Ahead-Funktion bei del.icio.us.

Hat man die Tags eingegeben und das Formular abgeschickt, leitet del.icio.us den Nutzer wieder zur gerade getaggten Webseite zurück, um ihm so die Möglichkeit zu geben, ohne lange Unterbrechung weiter zu surfen.

c Navigation in del.icio.us

Der Einstieg in del.icio.us erfolgt meist über die eigene Linksammlung, zu erreichen unter der URL http://del.icio.us/[Benutzername]/. Von dort aus kann man seine eigenen Bookmarks verwalten und in die gesamte Linksammlung aller del.icio.us-Nutzer einsteigen.

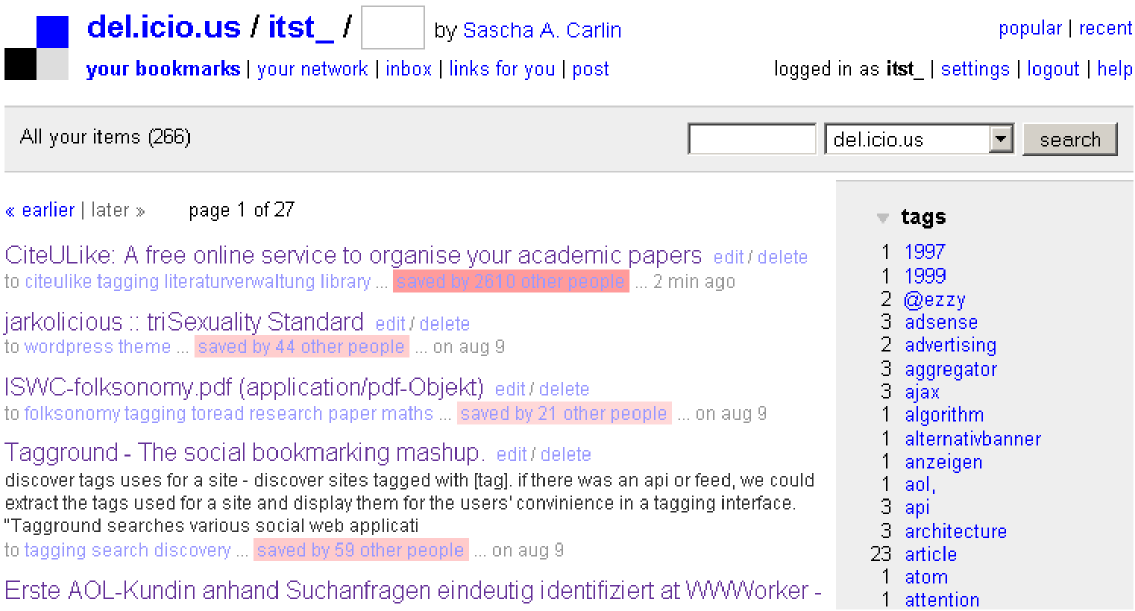


Abbildung 5-3: Ansicht der eigenen Linksammlung in del.icio.us.

Links zu sehen die letzten Bookmarks, rechts die Liste aller Tags, die man bereits benutzt hat. Über die Optionen kann man die Darstellung der Tags steuern und sie editieren.

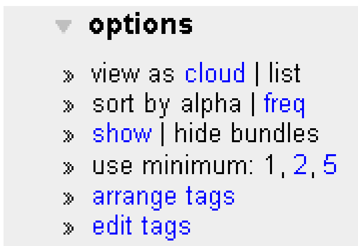


Abbildung 5-4: Optionen zur Darstellung der eigenen Tags.



Abbildung 5-5: Darstellung der eigenen Tags als Tag Cloud. Angezeigt werden nur Tags, die mindestens fünf Mal benutzt wurden.

So kann man die Tags nach dem Alphabet ordnen oder sie als Tag Cloud anzeigen lassen. Der Screenshot in Abbildung 5-5 zeigt die eigenen Tags in einer Tag Cloud, wobei nur die Tags, die man mindestens fünf Mal benutzt hat, angezeigt werden.

Wählt man aus der Liste nun ein Tag aus (im Beispiel hier „discussion“), so werden einem alle Ressourcen angezeigt, die mit diesem Tag versehen wurden. Außerdem erhält man eine Übersicht über die verwandten Tags, also jene, die man gemeinsam mit dem gewählten Tag benutzt hat.



Abbildung 5-6: Ausschnitt aus der Ansicht der eigenen mit "discussion" getaggten Ressourcen.

Aus der Liste der „related tags“ kann man nun ein weiteres Tag (hier: „collaboration“) auswählen und entweder zu den Ressourcen wechseln, die man mit ihm getaggt hat, oder man wählt das Plus-Zeichen vor dem Tag aus und erhält so eine Übersicht über die man mit „discussion“ und „collaboration“ getaggt hat. Wie bereits zu sehen ist, gibt es zwei Ressourcen, auf die das zutrifft.

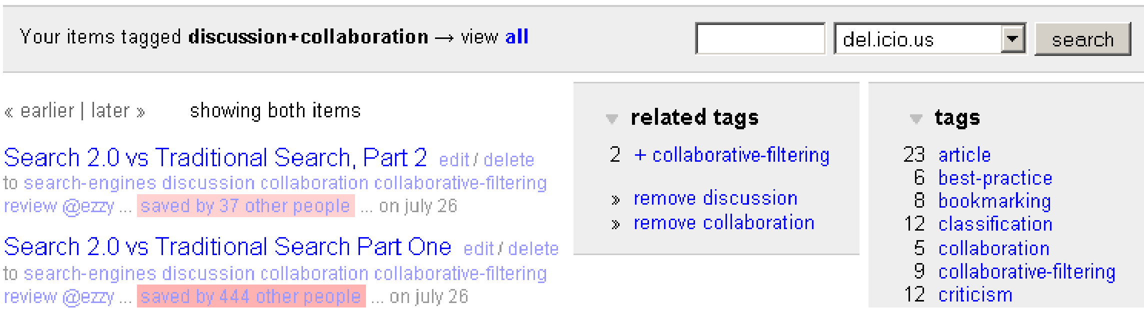


Abbildung 5-7: Ansicht der mit "discussion" und "collaboration" getaggten eigenen Ressourcen.

Über die beiden „remove“-Links kann man einzelne Tags aus dem Filter entfernen. Um nun zu sehen, welche Ressourcen andere Nutzer mit „discussion“ und „collaboration“ getaggt haben, wählt man in der Kopfzeile den Link „all“. In der nun angezeigten Liste findet man neben jeder Ressource einen Link „save this“, um diese Ressource seiner eigenen Sammlung hinzu zu fügen.

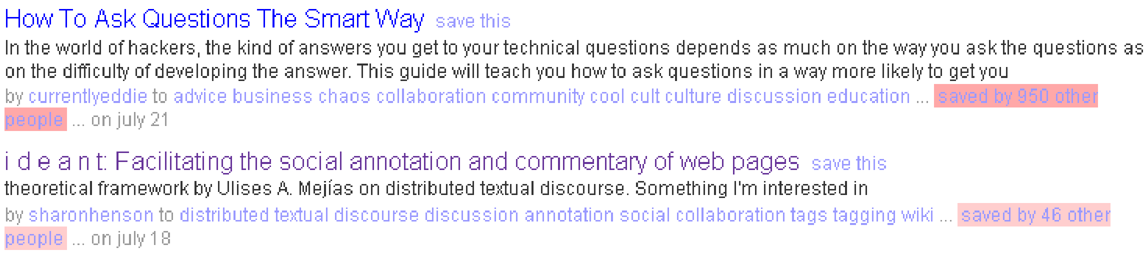


Abbildung 5-8: Ausschnitt aus der Liste aller mit "discussion" und "collaboration" getaggten Ressourcen.

Rot hinterlegt sieht man in der Darstellung die Anzahl der Personen, die diese Ressourcen bereits getaggt und ihrer Linksammlung hinzugefügt haben. Möchte man nun die Ressource „How To Ask Questions The Smart Way“ in seine Sammlung aufnehmen, wählt man „save this“ (rechts neben dem Titel) und gelangt zum Tagging-Formular:

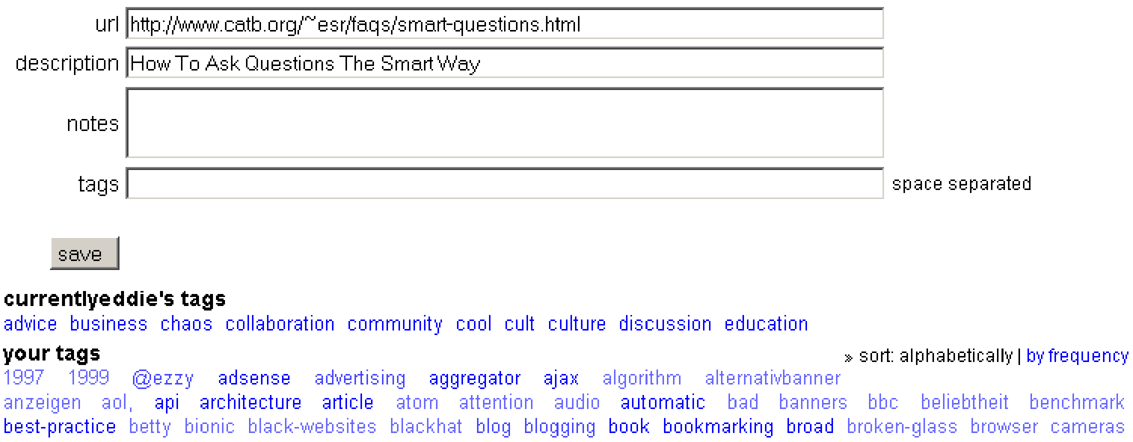


Abbildung 5-9: Ausschnitt aus dem Tagging-Formular mit den Tags, die von einem anderen Nutzer für diese Ressource vergeben wurden..

Nun werden neben den eigenen bereits genutzten Tags auch die Tags angezeigt, die „currentlyeddie“ benutzt hat, um diese Ressourcen zu taggen. „currentlyeddie“ ist der Nutzer, der diese Ressource zuletzt seiner Linksammlung hinzufügte, wie in Abbildung 5-8 zu erkennen.

Bisher haben wir uns nur in der eigenen Linksammlung sowie in einem globalen „Tag-Raum“ umgesehen. Auf die Art neue, interessante Ressourcen zu entdecken kostet viel Zeit und ist nicht praktikabel.

del.icio.us bietet jedoch einen Ausweg in Form von „subscriptions“. Diese können über ein einfaches Formular angelegt und verändert werden:



Abbildung 5-10: Die Benutzeroberfläche zum Anlegen und Editieren von Abonnements.

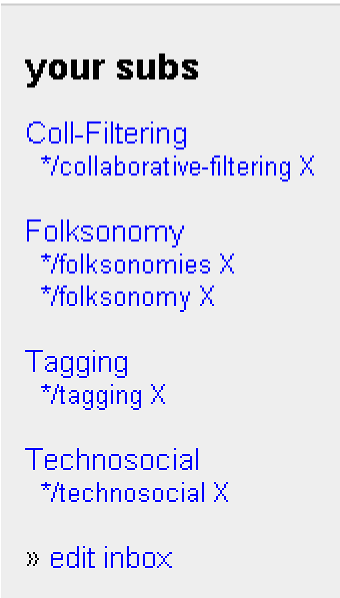


Abbildung 5-11: Darstellung der Abonnements in der "inbox". Abonnements mit dem gleichen "inbox name" werden in einerm "Ordner" zusammengefasst.

Diese Abonnements laufen nun in der „inbox“ (dritter Link von rechts in der zweiten Zeile) zusammen und können dort auf einen Blick betrachtet werden. Zu beachten ist die Funktion des „inbox name“. Die beiden Abonnements „*/folksonomies“ und „*/folksonomy“ (Der Stern zeigt an, dass diese Abonnements alle Ressourcen mit diesen Tags umfassen, gleich welcher Benutzer sie getaggt hat. Wollte man nur die Ressourcen eines bestimmten Benutzers abonnieren, würde dies so aussehen: „[Benutzername]/[Tag]“.) haben beide den gleichen „inbox name“. Sie erscheinen in der „inbox“ gemeinsam.

Darüber hinaus kann man auch einen Benutzer beobachten und so regelmäßig sehen, welche Ressourcen er neu getaggt hat. del.icio.us nennt dieses Feature „network“.

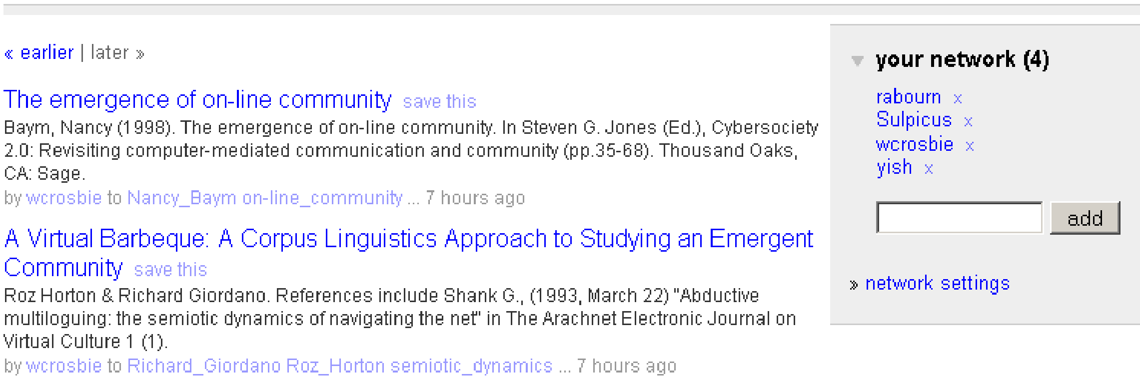


Abbildung 5-12: Ausschnitt aus der Darstellung des "network" eines Nutzers.

In diesem (siehe Abbildung 5-12) Netzwerk sind die neu getaggten Ressourcen der Benutzer „rabourn“, „Sulpicus“, „wcrosbie“ und „yish“ zusammengefasst.

Eines der interessantesten Features von del.icio.us ist aber die Möglichkeit, für nahezu alle möglichen Kombinationen von Benutzern und Tags die jeweils neuesten Ressourcen per RSS-Feed abzurufen und entweder direkt im Browser (bei Firefox „Live-Bookmarks“ genannt“) oder mittels einem Feed-Reader zu lesen. Jede Seite, auf der eine Liste Ressourcen angezeigt wird, kann so „exportiert“ und auf anderen Webseiten oder in Feed-Readern dargestellt werden.

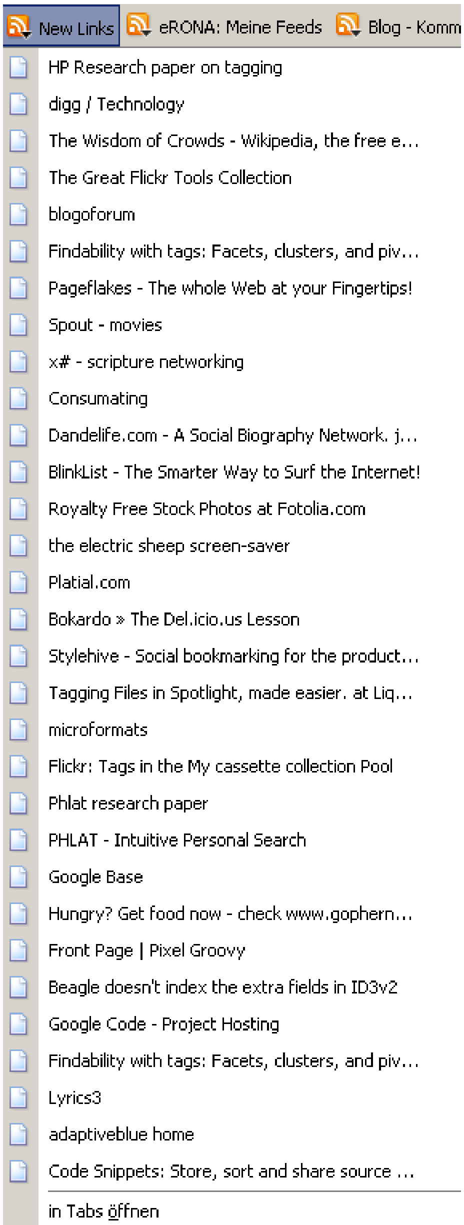


Abbildung 5-13: Ansicht des RSS-Feeds einer "inbox" im Browser Firefox.

Eine der vielen Anwendungsmöglichkeiten ist die Kombination der Feeds mit der im Abschnitt Filtern statt suchen, S. 73f, erwähnten Gruppentags. Hat sich eine Gruppe auf ein bestimmtes Tag geeinigt, das sie benutzen will, um alle Mitglieder der Gruppe stets über neue Ressourcen zu informieren, kann man die so getaggten Ressourcen auf einer externen Webseite zusammenfassen. del.icio.us sorgt dafür, dass der Feed regelmäßig aktualisiert wird.

6 Tagging als Chance

a Informationsflut

Die Entwicklung des World Wide Web und der Möglichkeiten des Auffindens relevanter und interessanter Ressourcen zeigt, dass sowohl Kataloge wie Volltextsuchmaschinen nur zum Teil in der Lage sind, mit dem exponentiellen Wachstum des Webs, dass heißt der Websites und deren Inhalten, Schritt zu halten. Kataloge sind aus dem Suchprozess der meisten Nutzer vollständig verschwunden – zumeist werden Volltextsuchmaschinen angefragt, in Deutschland vor allem Google. Laut den Daten von Webhits[43] liegt dabei Google mit 86% Prozent aller Suchanfragen an Platz eins. Der erste Katalog, AllesKlar, taucht auf Platz neun auf mit lediglich 0,3%. Der DMOZ-Katalog, einer der früher führenden Kataloge, belegt lediglich Platz 14 mit 0,1% der Suchanfragen. In den USA erreichen Google, Yahoo und MSN gemeinsam ebenfalls 86%[44] und lassen damit andere Suchmaschinen und Kataloge weit hinter sich.



Abbildung 6-1: Screenshot eines Artikel von heise.de. Neben dem eigentlichen Inhalt in der Mitte indexieren Volltextsuchmaschinen auch die Inhalte rechts und links davon.

43 Siehe http://www.webhits.de/deutsch/index.shtml?/deutsch/webstats.html – Zugegriffen am 10.08.2006. Stand vom 10.08.2006.
44 Siehe http://www.comscore.com/press/release.asp?press=935 – Zugegriffen am 10.08.2006. Stand vom 18.07.2006.

Volltextsuchmaschinen im World Wide Web haben allerdings gegenüber Textdatenbanken wie denen von Genios und anderen Informationsdiensten den Nachteil, über wesentlich weniger Suchoperatoren zu verfügen. Zwar beherrscht der Marktführer Google boolesche Operatoren und Synonymsuche, aber es lässt sich beispielsweise nicht nach Begriffen suchen, die im gleichen Satz oder Abschnitt vorkommen. Auch indexieren Google, Yahoo und MSN die gesamte Seite, inklusive aller Daten, die zum Layout und der Navigation einer Site gehören, anstatt nur den eigentlichen Inhalt der Seite zu indexieren.



Abbildung 6-2: Google findet für die angegebenen Suchbegriffe, die aus der Navigation von heise.de stammen, über vier Millionen Treffer allein auf heise.de selbst.

Volltextsuchmaschinen müssen, wenn sie auf Dauer im World Wide Web bestehen wollen, nicht nur ihre Suchalgorithmen ständig verbessern, sondern auch Wege finden, irrelevante Informationen aufzufiltern.

b Filtern statt suchen

Taggingsysteme verfolgen einen anderen Ansatz. Statt Volltextsuchen bieten sie Zugriff auf Informationen, die von Menschen indexiert, getaggt, wurden. Wie gezeigt wurde, ermöglichen sie es so, relevante Inhalte zu entdecken und diejenigen Inhalte herauszufiltern, die für den Suchenden von Interesse sind. Wenn die ersten Ansätze zu Collaborative Filtering in Taggingsystemen weiter ausgebaut werden, bieten sich hier interessante Möglichkeiten, sehr einfach und schnell Inhalte zu entdecken, auf die man sonst wahrscheinlich nicht gestoßen wäre. Kein einzelner Mensch hat die Ressourcen oder die Möglichkeit, aus den über acht Milliarden Webseiten diejenigen herauszufinden, die für ihn interessant wären. Doch durch die Masse an Menschen, die Taggingsysteme wie del.icio.us benutzt, hat er die Möglichkeit, auf das Wissen – dass heißt auf die Informationen, die alle Nutzer gemeinsam eingegeben haben – aller zuzugreifen.

Dienste wie del.icio.us bieten kleineren Gruppen von Menschen die Möglichkeit, sich gegenseitig auf interessante Informationen aufmerksam zu machen. Beispielsweise durch die Benutzung eines speziellen Gruppentags erhält man einen Gesamtüberblick über alle Ressourcen einer Gruppe. Durch Kombinierung des Gruppentags mit anderen Tags filtert man so das Suchergebnis immer weiter, bis man schließlich diejenigen Ressourcen entdeckt hat, die im Moment relevant sind.

7 Ausblick

a Kombination von Volltextsuchmaschinen und Tagging

Mit dem Kauf von del.icio.us und Flickr hat Yahoo gezeigt, dass man dort an diese Systeme glaubt. Yahoo möchte die del.icio.us-Plattform schnellstmöglich in das eigene Produkt MyWeb2 integrieren. Damit könnten Yahoo-Nutzer beide Welten, die der Volltextsuchmaschinen und der Taggingsysteme, parallel und gemeinsam nutzen. Sucht man eine spezielle Information, nutzt man die Volltextsuchmaschinen. Möchte man weitere Inhalte entdecken, nutzt man die Tags und die Folksonomy.

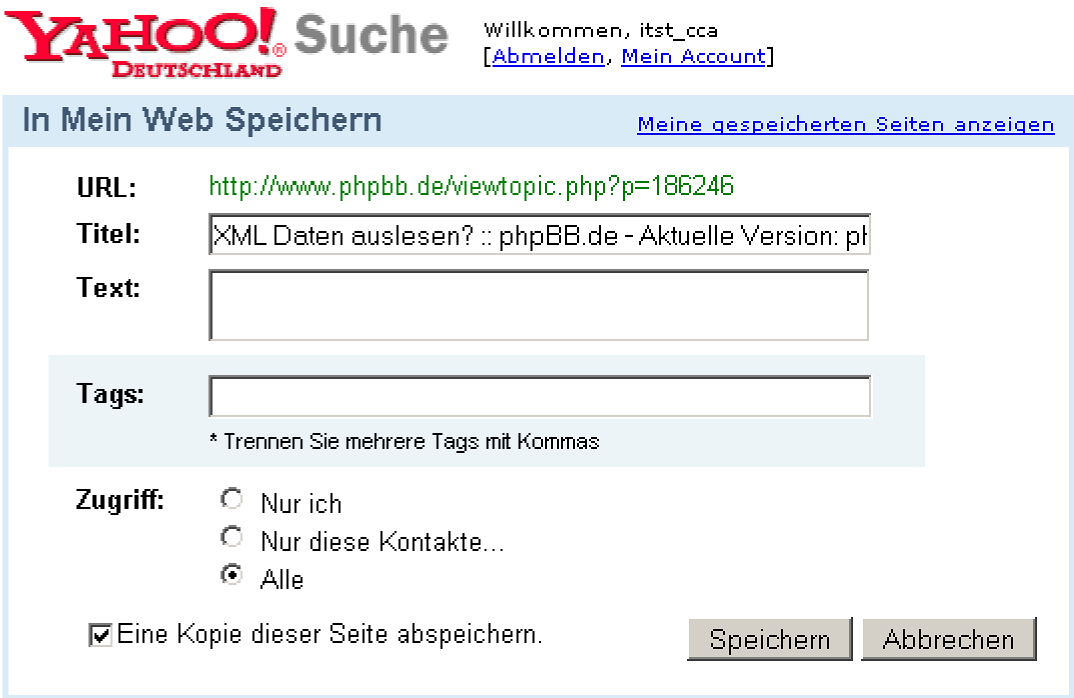


Abbildung 7-1: Eine mittels der Volltextsuche gefundene Ressource kann der Nutzer in seinem "Mein Web" ablegen und dazu Tags vergeben.

Über seine eigenen Tags gelangt der Nutzer zu allen Ressourcen, die ebenfalls mit diesem Tag getaggt wurden. Hier hat er die Möglichkeit, tiefer in die Ergebnisse einzusteigen und weitere Inhalte zu entdecken.



Abbildung 7-2: Ausgehend von einem selbst vergebenen Tag (hier "tags") gelangt der Nutzer zu einer Liste aller damit getaggten Ressourcen und hat die Möglichkeit, "tags" mit weiteren Tags zu kombinieren und damit die Ressourcen weiter zu filtern.

In Abbildung 7-3 ist der Nutzer, ausgehend von „tags", tiefer in die getaggten Ressourcen eingestiegen und hat „tags" zunächst mit „folksonomy" und schließlich mit „web2.0" kombiniert, um die Trefferliste zu filtern.

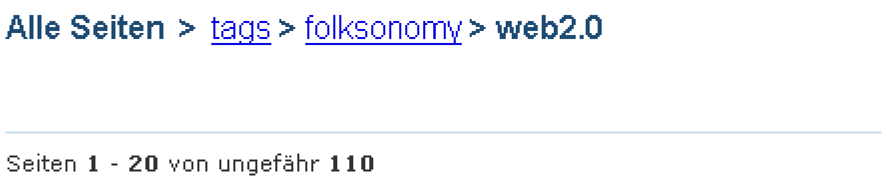


Abbildung 7-3: Durch Filterung nach weiteren Tags hat sich die Treffermenge auf 110 reduziert.

b Tagging im Unternehmen

IBM hat mit Dogear[45] in einer Eigenentwicklung nicht nur das Prinzip von del.icio.us übernommen, sondern es mit Erweiterungen versehen, die in einem großen Konzern notwendig und sinnvoll sind. So kann der Nutzer nicht nur Anmerkungen zu seinen Ressourcen hinzufügen, sondern er kann auch per „Alerts“ auf Inhalte aufmerksam gemacht werden, die in seiner Interessensphäre liegen. Dogear bietet auch Ansätze von Collaborative Filtering, um Nutzer auf Inhalte aufmerksam zu machen:

> *Dogear also exploits collaborative filtering techniques to screen new bookmarks for those that are predictably of interest to an individual (or a group of individuals). Common interests can be inferred based on a number of observable user actions, including use of similar tags and/or tag combinations, similarity of bookmark (URL) collections, common RSS subscriptions, and click streams that indicate interest in specific kinds of bookmarks. Text analysis of bookmark titles, descriptions, and comments will also be used to determine bookmark relatedness. E-mail notification, system tray status indicators, or small pop-up windows can alert users to new information.*

Dogear wertet dabei nicht nur die von einem Benutzer vergebenen Tags aus, sondern betrachtet auch die Verbindungen zwischen Benutzern und Ressourcen sowie die abonnierten Feeds und die angeklickten Bookmarks.

45 Siehe http://domino.research.ibm.com/comm/wwwr_thinkresearch.nsf/pages/20060627_dogear.html – Zugegriffen am 10.08.2006.

IBM wird Dogear, das momentan im Intranet von IBM läuft, in die kommende Version von Lotus Notes, Hannover, integrieren[46] und damit die Idee des Taggens in all die Unternehmen tragen, die Lotus Notes einsetzen. Weltweit sind dies 60.000 Unternehmen mit insgesamt 120 Millionen Nutzern[47].

c Amazon Tags

Amazon bietet auf seiner US-amerikanischen Website Nutzern die Möglichkeit, Produkte zu taggen und davon ausgehend, weitere Produkte zu entdecken. Dieses Feature steht parallel zu den Empfehlungen und den Wunschzetteln, die Amazon darüber hinaus anbietet. Eine Integration aller drei Systeme wird angestrebt und könnte die Empfehlungen, die momentan auf tatsächlich getätigten Käufen basiert, verbessern[48]:

> *You might want to tag items you've purchased in the past. This allows you to organize your collection in any way you wish. You can also use tags to help you remember items your're considering for purchase, e.g. if you are planning to buy a digital camera, you might tag several makes and models so that you can look at them side by side, organized the way you wish. Don't like the way Amazon has organized its science fiction or cookbooks? Tag the items you care about with the categories that matter to you. Because people's tags are (by default) visible to others, a great side effect of tagging is that you can navigate among items through other people's tags. What items have people tagged "gift" or "Tuscany" or "robot?"*

d Aus Folksonomies hierarchische Strukturen bilden

Heymann und Garcia-Molina (vgl. [Heymann2006]) schlagen eine Methode vor, um eine flache Folksonomy in eine Hierarchie umzuwandeln. Sie gehen dabei von einem triparaten

46 Siehe http://www.computerwoche.de/produkte_technik/software/576713/ - Zugegriffen am 10.08.2006. Stand 24.05.2006.
47 Telefonische Information der deutschen IBM-Pressestelle vom 10.08.2006.
48 Siehe forums.prosperotechnologies.com/am-custreview/messages?msg=19295.1 – Zugegriffen am 10.08.2006.

Graphen aus und betrachten die Häufigkeit, mit der ein Tag vorkommt und die weiteren Tags, die häufig mit diesem Tag vorkommen.

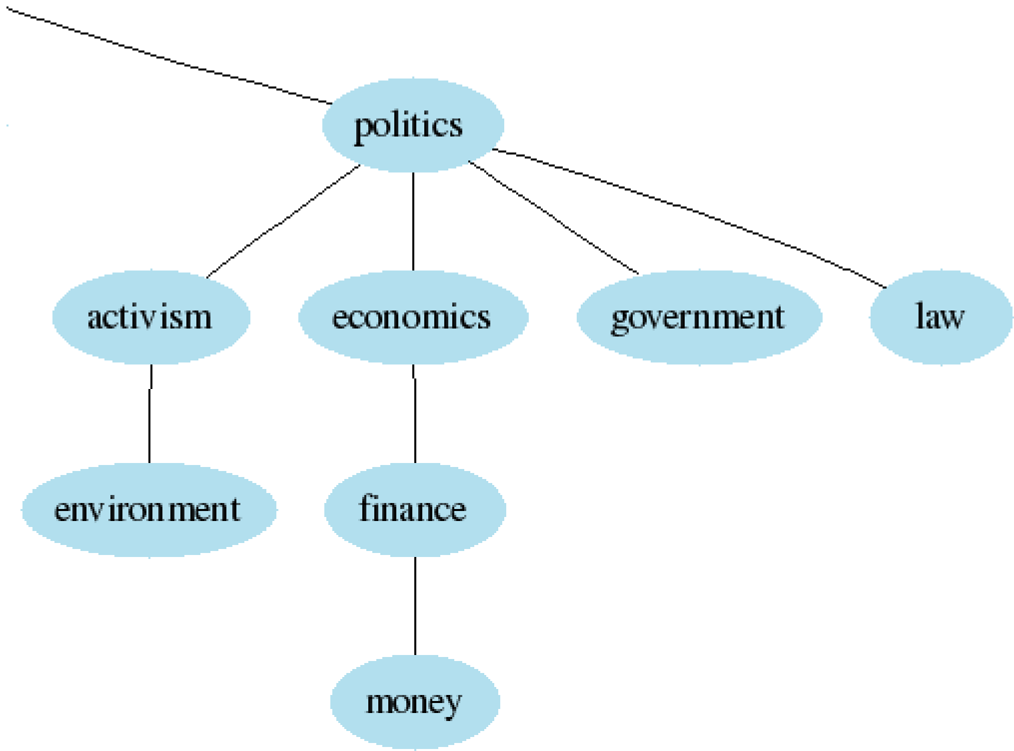
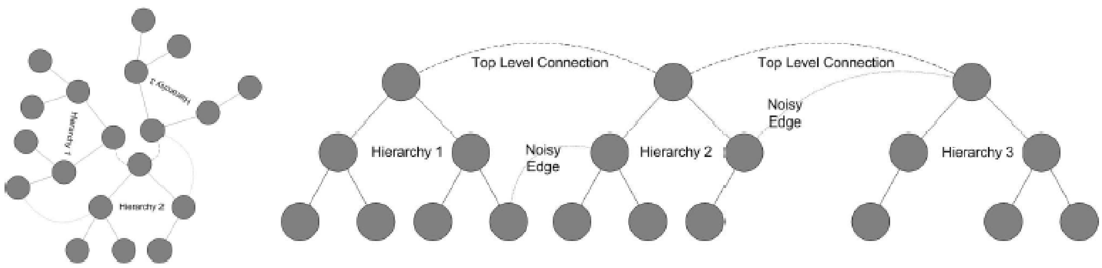


Abbildung 7-4: Ausschnitt aus einer Tag-Hierarchie, erstellt nach der Methode von Heymann und Garcia-Molina.

Diese Ähnlichkeiten überführen sie in einen ungewichteten Graphen und kappen dabei diejenigen Verbindungen zwischen zwei Tags, bei denen die Ähnlichkeit unter einem gesetzten Schwellenwert liegt. Nun bewegen sie sich durch diesen Graphen und fügen Tag für Tag zu einem neuen Graphen hinzu. Dabei ordnen sie jedes Tag jeweils demjenigen Tag unter, mit dem es die meiste Ähnlichkeit hat. Hat ein Tag eine kleinere Ähnlichkeit mit allen bereits im Graph vorhandenen Tags, wird es der Wurzel des Graphen untergeordnet; es ist scheinbar ein neuer übergeordnetes Tag.

Heymann und Garcia-Molina stellen fest, dass ihre Methode umso besser funktioniert, je größer die zugrunde liegende Folksonomy ist. Sie prüften ihren Algorithmus anhand von Daten aus del.icio.us und CiteULike und bemerken dazu:

With respect to most of these features, the CiteULike data turns out to be more difficult to turn into a hierarchy: it has low density (some users do not tag objects at all), low overlap between users (academics working in different fields), and a specificity distribution which is very highly geared towards detailed tags. By contrast, Delicious is high density, high overlap, and has a much more even specificity distribution. While different algorithms might be better tuned for different specificity distributions, low density and overlap make it generally more difficult to create a hierarchy from tagging data (these factors mean that our algorithm needs a larger sample dataset from CiteULike than from Delicious to be effective).



(a) Similarity Graph　　(b) Latent Hierarchical Taxonomy

Abbildung 7-5: Die beiden Graphen, die Heymann und Garcia-Molina nutzen, um aus einer Folksonomy eine Hierarchie zu generieren. Rechts der ungewichtete Ähnlickeitsgraph, links die sich ergebende Hierarchie.

e Tag-Cluster

Begelman, Keller und Smadja (vgl. [Begelman2006]) verfolgen einen anderen Ansatz. Sie versuchen Tag-Cluster zu bilden und beschreiben dazu eine Vorgehensweise, die von der Häufigkeit ausgeht, mit der ein Tag a mit einem anderen Tag b in derselben Tag-Gruppe vorkommt. Anhand von Daten von del.icio.us identifizierten sie aus einer Auswahl von Tag-Gruppen 1100 Tag-Verbindungen, der sie jeweils eine Häufigkeit zuordneten.

Sie überführen nun diese Datentabellen in einen Funktionsgraphen und bilden die erste und die zweite Ableitung der Funktion. Sie suchen nun, beginnend am Endes des Graphen, nach dem Punkt, an dem die erste Ableitung ein lokales Maximum erreicht bzw. die zweite Ableitung vom Positiven ins Negative übergeht. An diesem Punkt bestimmen sie den „cut-

tag	*count*	*tag*	*count*
feed	310	web2.0	77
blog	298	home	65
feeds	246	wikipedia	59
search	219	blogs	57
news	173	biography	53
google	103	preview	48
xml	102	learn	33
web	81	sitemap	30

Abbildung 7-6: "Co-Tags" von "rss" und die Häufigkeit, mit der diese Tags gemeinsam mit "rss" in einer Tag-Gruppe vorkommen.

off“: alle Tags rechts davon werden als nicht verwandt genug verworfen, die übrigen links davon als stark verwandt angesehen. Dabei bemerken sie, dass der einzige zu optimierende Parameter das „minimal peak high“ ist. Läge dieser Wert zu niedrig, würde bereits ein kleiner Peak als „cut-off“ genutzt werden. Wenn die erste Ableitung einer Verteilung kein Maximum größer als der angenommene Wert von „minimal peak high“ hat, wird immer das Tag mit der größten Häufigkeit als stark verwandt angesehen und der Rest verworfen. So identifizieren sie Cluster von Tags, die sich so ähnlich wie möglich sind.

- Query tag: *health*:
  - shopping, research
  - nutrition, food, diet
  - fitness, workout, running
  - article, science
  - life, lifehack, product,howto, gtd, reference, tip
  - esport, sport
- Query tag: *sports*
  - hockey, nhl
  - basebal, mlb, triple
  - basketbal, nba, nbdl, wnba
  - footbal, nfl
  - alcohol, beer, tv, food, bar
  - computer game, action game, free game

Abbildung 7-7: Cluster für "health" und "sports", gewonnen nach der Methode von Begelman, Keller und Smadja.

Begelman, Keller und Smadja beschreiben ihren Algorithmus als so schnell, dass es vorstellbar wäre, die Cluster „on-the-fly“, also während des Zugriff, zu generieren.

8 Fazit

Die Forschung an Tagging und Folksonomy steht noch am Anfang. Doch die bereits publizierten Arbeiten lassen darauf schließen, dass Tagging durchaus noch längere Zeit ein interessantes Thema bleiben wird. Tagging kann und sollte klassische Indexierungsverfahren und Klassifikationen weder ersetzen noch ablösen. Es ist als Komplement zu den bestehenden Volltextsuchmaschinen auf der einen und strikt reglementierten Katalogen auf der anderen Seite zu sehen.

Dass Collaborative Filtering funktioniert, zeigt nicht nur Amazon mit seinen Empfehlungen. Für Anwendungen auf Basis von Collaborative Filtering, stehen mit den durch Tagging erzeugten Folksonomies Datenbasen bereit, mit denen, besser als Buchkäufe das können, genauer und zielgerichteter gefiltert werden kann. Das erläuterte Beispiel einer Collaborative Filtering-Anwendung für CiteULike ist ein Beispiel dafür.

Mit dem Wissen über Clustering-Algorithmen und den Möglichkeiten der verschiedenen Graphendarstellungen von Folksonomies ist es kein weiter Weg mehr zu quasi-hierarchischen Folksonomy-Abbildungen. Diese Anwendungen müssen natürlich der Natur der Folksonomies, sich regelmäßig zu ändern, Rechnung tragen. Die Bildung von Clustern könnte sich als Weg erweisen, dieses Problem zu lösen.

9 Anhänge

a Abbildungsverzeichnis

b Tabellenverzeichnis

c Literaturverzeichnis

[Anderson2004]
Anderson, Chris: *The Long Tail*. Wired Magazine, Vol. 12, Issue 10. Oktober 2004. http://www.wired.com/wired/archive/12.10/tail_pr.html – Zugegriffen am 07.08.2006.

[Begelman2006]
Begelman, Grigory; Keller, Philipp & Smadja, Frank: *Automated Tag Clustering: Improving search and exploration in the tag space*. In the Proceedings of WWW 2006 Collaborative Web. Tagging Workshop, Edinburgh, Scotland. 2006. http://www.rawsugar.com/www2006/automatedTagClustering.pdf – Zugegriffen am 08.08.2006.

[Chun2006]
Chun, Susan et al.: s*teve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums*. Proceedings Muesums and the Web 2006: http://www.archimuse.com/mw2006/papers/wyman/wyman.html – Zugegriffen am 06.08.2006.

[Cialdini2001]
Cialdini, R.: *Influenece – Scienece and Practice.* Allyn & Bacon. 4. Auflage, 2001. Zitiert nach [Golder2005].

[Dic1983]
Dictionary of Contemporary English. Langenscheid-Longman, 4. Auflage, 1983.

[Fugmann1999]
Fugmann, Robert: *Inhaltserschließung durch Indexieren: Prinzipien und Praxis*. Deutsche Gesellschaft für Dokumentation e. V., Frankfurt am Main. 1999.

[Golder2005]
Golder, Scott & Huberman, Bernado A.: *The Structure of Collaborative Tagging Systems*. Information Dynamics Lab, HP Labs. 18.08.2005:
http://arxiv.org/abs/cs.DL/0508082. Zugegriffen am 03.06.2006

[Guy2006]
Guy, Marieke & Tonkin, Emma: *Folksonomies: Tidying up Tags?*. D-Lib Magazine, Vol. 12, Issue 1. Januar 2006:
http://www.dlib.org/dlib/january06/guy/01guy.html – Zugegriffen am 25.06.2006.

[Hammond2006]
Hammond, Tony et al.: *Social Bookmarking Tools (I)*. D-Lib Magazine, Vol. 11, Issue 4. April 2005:
http://www.dlib.org/dlib/april05/hammond/04hammond.html – Zugegriffen am 05.07.2006.

[Heymann2006]
Heymann, Paul & Garcia-Molina, Hector: *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems.* Department of Computer Science, Stanford University, Stanford, CA, USA, April 2006.
http://dbpubs.stanford.edu:8090/pub/2006-10 – Zugegriffen am 10.08.2006.

[Kress2003]
Kress, Gunther: *Literacy in the new media age*. Routledge, 2003. Zitiert nach [Mejas2005].

[Lambiotte2005]
Lambiotte, R. & Ausloos, M.: *Collaborative tagging as a tripartite network*. SUPRATECS, Unversité de Liège, Liège, Belgium. 09.07.2005.
http://arxiv.org/abs/cs.DS/0512090 – Zugegriffen am 10.07.2006.

[Marlow2006]
Marlow, Cameron et al.: *Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead*. In the Proceedings of WWW 2006 Collaborative Web. Tagging Workshop, Edinburgh, Scotland. 2006.
http://www.rawsugar.com/www2006/29.pdf – Zugegriffen am 15.05.2006.

[Mejias2005]
Mejias, Ulises Ali: *Tag Literacy*. 26.04.2005.
http://ideant.typepad.com/ideant/2005/04/tag_literacy.html – Zugegriffen am 20.07.2006.

[Melzer2005]
Melzer, James: *The best of del.icio.us*. 16.11.2005
http://www.jamesmelzer.com/bearings/archives/2005/11/the_best_of_del.html – Zugegriffen am 08.08.2006.

[Melzer2006]
Melzer, James: *Making social bookmarking better*. 19.02.2006. http://www.jamesmelzer.com/bearings/archives/2006/02/making_social_bookmark.html – Zugegriffen am 08.08.2006.

[Mika2005]
Mika, Peter: *Ontologies are us: A unified model of social networks and semantics*. In Proceedings of the 4th International Semantic Web Conference, 2005. http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf – Zugegriffen am 08.08.2006.

[Norton2006]
Norton, Quinn: *'I want to build someting that grows'*. Guardian Unlimited | Technology, 26.01.2006: http://technology.guardian.co.uk/weekly/story/0,,1694487,00.html – Zugegriffen am 19.07.2006.

[Sinah2005]
Sinah, Rashmi: *A cognitive analysis of tagging*. 27.09.2005: http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html – Zugegriffen am 16.07.2006.

[Torres2005]
Torres, Elias: *Why I don't like to infer semantics from tags*. 21.07.2006. http://torrez.us/archives/2006/07/21/473/ - Zugegriffen am 09.08.2006.

[VanderWal2005]
Vander Wal, Thomas: *Explaining and Showing Broad and Narrow Folksonomies.* 21. Februar 2005: http://www.vanderwal.net/random/entrysel.php?blog=1635 – Zugegriffen am 06.07.2006.